A Random Forests based Performance Ratio for Regulatory Asset Portfolio Management and Optimization

Boris Waelchli
Department of Banking and Finance, University of Zurich,Plattenstrasse 14
CH-8032 Zuerich, Switzerland
boris.waelchli@uzh.ch

November 4, 2015

Abstract

The following paper proposes a portfolio performance measure to optimize, mostly bond asset portfolios usually held for regulatory purposes from a risk focused perspective. The measure is based on variations of the proximity measure introduced by the Random Forests framework, leading to a proximity based performance ratio. The proximities are modeled using a recursive conditional partitioning type of Random Forests, which allows for a ranking as well as an analysis of the risk drivers of the portfolio performance. The proximity based performance ratio is shown to, on average, outperform nine different and commonly known risk and performance ratios as well as the 1/N-balanced portfolio in three different tests, in- and out of the sample. The proximity based performance ratio can consider a large amount of risk rivers and is suitable for big data analysis for big and small financial institutions.

Keywords: Random Forests; Recursive Conditional Partitioning; Regulatory Portfolio Optimization; Sharpe Ratio; Big Data
JEL: C02, D81, G11, G32

1 Introduction

As a consequence of the financial crisis in 2007 and the resulting regulatory requirements, the demand for high quality assets is rising. Subsequently the portfolios banks of all sizes are holding for regulatory purposes are increasing in volume. The majority of these portfolios contain a large share of government bonds. Bonner (2014) have in their work identified that most banks are indeed acquiring large amounts of government bonds even beyond their risk appetite for such products. The intention of holding assets for regulatory purposes is risk management, at least in the eyes of the regulator.
The increased number of such portfolios is thus characterized by a risk focused management, while their composition is mostly stable and made of government bonds, short selling is not allowed. Smaller financial institutions portfolios especially qualify as being sparse, which means that they are stable with few actively restructured positions.

However, empirically, almost all portfolios' returns are not normally distributed but are negatively skewed and exhibit an elevated or high kurtosis, in other words: fat tails.

In the market, the classic Sharpe ratio is still the most commonly and widely used performance measure for portfolios. This despite the fact that it requires portfolio returns to be independent and identically normal distributed. This is in contrast to the empirically observed negative skewness and high positive excess kurtosis. Furthermore the returns often exhibit serial auto-correlation and heteroskedasticity, which calls the independence and identity into question. These are well known issues in the literature. Cogneau and Hubner (2009) have identified and discussed more than 100 portfolio performance measures within the current literature. Most of these are proposed amendments to the classic Sharpe ratio trying to solve the issues mentioned above. Consequently alternative measures often aim at higher moments of the distribution to get rid of assumptions on the distribution at all. Cogneau and Huebner, as well as other authors in the literature in general, find that there is no conclusion as to which performance measure is best, but a majority of scholars state that the amendments of the Sharpe ratio add information to the portfolio optimization process.

This paper provides a utility based portfolio optimization methodology tailored to regulatory asset portfolios. The performance is measured using a risk return ratio similar to the Sharpe ratio with the proposed approach being based on a variant of Random Forests proximities; allowing to use the advantages of the Random Forests algorithm. Random Forests can handle very large amounts of data, which renders it suitable for institutions with long and extensive data histories and big data applications. On the other hand, as will be shown in the analysis section, the approach is especially suitable for smaller institutions as good results can be achieved by application of a large amount of external and publicly available risk indicators. The Random Forests proximity based approach does not require assumptions on the distribution of the included risk indicators, or the underlying returns. It can assure stability of the results due to large amounts of trees. Correspondingly the method needs less historical data in comparison to volatility based models. The modeling of the proximities by risk indicators allows the linking of risk contribution concepts to the underlying risk drivers: by using Random Forests importance measures when modeling the proximities, the risk drivers which mostly contribute to the loss classification and implicitly to the proximities, can be identified. This allows a ranking as well as an analysis of the risk drivers for additional risk and performance management purposes.

Overall the model is shown to outperform a set of benchmark risk-return ratios in bond portfolio performance, while being itself mathematically relatively simple.

Relevant Literature about Portfolio Optimization

This paper contributes to the literature on portfolio optimization in four ways.

Firstly, in the optimization of skewed, fat tailed and sparse portfolios. Cogneau and Hubner (2009) have analyzed 101 alternatives and found that no methodology clearly outperforms the others. This paper will benchmark against a choice of alternatives cited in Cogneau and Huebner and show that the proximity based ratio outperforms them in most tested cases in- as well as out of the sample. DeMiguel et al. (2009b) have shown in a similar attempt that none of the alternative portfolio strategies within their paper significantly outperforms the balanced or 1/N portfolio strategy. This paper includes the 1/N strategy as an alternative as well. The portfolio performances using alternative strategies are tested statistically for significant differences. Analog to DeMiguel et al. (2009a) who test for significant differences of strategies on portfolios with constrained norms, the statistical test applied in this paper is the two sided studentized circular block bootstrap test developed by Ledoit and Wolf (2008) for non-normal distributions.

Secondly, the proposed measure encourages sparse portfolios, following recent studies like Brodie et al. (2008) who show that more stable, sparse portfolios can be constructed by application of constrained linear regressions and the omission of short selling in the portfolio. Fastrich et al. (2014) provide a framework to construct sparse portfolios by regularization methods that aim at defining a penalty by controlling the fitting of the asset weight vector to the estimation errors in the covariance matrix. In this paper the sparse portfolio is encouraged by the omission of short selling, an exposure limit in the form of a maximum weight per asset and by the proximity based ratio itself. The later will be shown to encourage sparsity by comparison of the turnover rate, as applied in DeMiguel et al. (2009b), to the turnover rate of the alternative strategies.

Thirdly, this paper investigates the sources of risk in an optimized portfolio and the risk contributions of the assets as well as of the risk indicators. Recent papers that consider the contribution of risk indicators are Tasche (2008) who uses the Euler allocation to calculate contributions of individual names to CDO tranche losses. He additionally proposes a measures for the impact of risk factors in the non-linear case. Cherny and Madan (2007) likewise study the risk contribution of individual positions given the risk factors. Iscoe et al. (1999) and Iscoe and Kreinin (2000) focus on the impact of systematic a well as idiosyncratic risk factors of individual counterparties in structural models. This paper contributes by using the importance measure derived from the influence of each risk factor in fitting a Random Forests model to the performance of assets. With the importance measure, risk factors can be linked to the outcome of the proximity based performance ratios for individual assets.

Fourthly, basing portfolio optimization on Random Forests proximities encourages the usage of big data, since Random Forests allows the inclusion of a relatively unlimited amount of risk indicators. This happens without overfitting as long as some of the risk indicators are more than noise and actually add information to the classification (Biau 2012). In comparison, Jothimani et al. (2014) have developed a 5 stage methodology to apply big data to asset selection and weighting. The stages include *shortlisting stocks using data envelopment analysis, incorporation of the qualitative factors using text mining, stock clustering, stock ranking and optimizing the portfolio using optimization heuristics*, (Jothimani et al. 2014). This paper adds to the current literature by proposing a ratio that is able to accommodate big data in relation to portfolio optimization in a mathematically simple and straightforward way.

The remainder of this paper will be organized as follows: Section 2 introduces the technical background of the applied Random Forests algorithm and develops the variants of the proximities and the proximity based risk return ratio. It also introduces alternative benchmark strategies. Section 3 is concerned with the application of the portfolio optimization to a scenario simulation, a Monte Carlo simulation and an empirical study. This section further includes the analysis of the results and statistical tests for the significance of the deviation within the performance of all applied strategies. Section 4 concludes the paper.

2 Proximity Based Portfolio Model Approach

A performance ratio is constructed using the variations of the Random Forests proximities defined below: the downside- and upside proximities as well as the proximity based hedge effect to optimize the portfolio composition. Because the proximities result from a Random Forests models built on risk indicators, the resulting proximities directly link the performance and optimization of the portfolio with these risk drivers. Using Random Forests, an importance measure for each included risk indicator can be calculated and thus the most important risk drivers can be identified.

Nonetheless, as pointed out by several scholars (for example (Strobl et al. 2007)) the classic Random Forests is not the best choice to assess the influence of individual variables on the results. It has repeatedly been shown to prefer continuous splitting variables or variables with many different realizations to variables with only a few realizations. Consequently, with regard to the identification of influential risk drivers the conditional recursive partitioning framework, developed in Hothorn et al. (2006a) and Strobl et al. (2007) is applied instead of the classic Random Forests by Breiman. This is because the conditional recursive partitioning framework is unbiased in the choice of splitting variables and thus allows for unbiased variable importance measures and interpretation of the model as such.

2.1 Theoretical background of Recursive Conditional Partitioning

Overall, the conditional recursive partitioning forest is very similar to the classic Random Forests. The basic difference is not selecting variables for splitting based on the optimization of a partition of a two dimensional space but to choose the variable which has the highest association to the dependent variable based on a linear statistic.

To describe the framework in detail, the notation is introduced first: Assuming a dataset with n independent variables $\mathbf{X} := (\mathbf{X}_1, ..., \mathbf{X}_n)$, with $\mathbf{X}_i := (X_{1i}, ..., X_{mi})$, one dependent variable $\mathbf{Y}$, with $\mathbf{Y} := (Y_1, ..., Y_m)$ and m observations $\mathcal{O}_j := (Y_j, \mathbf{X}_j)$ for $j \in \{1, ..., m\}$, while $\mathcal{O} := (\mathbf{Y}, \mathbf{X})$. For this purpose, $\mathbf{Y}$ is binary and describes the outcome non-event/event thus $Y_j \in \{0, 1\}$. Thus $\mathbf{X}_i :\in \mathbb{R}^m$, $\mathbf{Y} \in \{0, 1\}^m$ and $\mathcal{o}_i \in (y_j, \mathbf{x}_j)$. The forest then works in the following way:

1. For each tree a training sample of a predefined size $\mathcal{O}_s := (\mathbf{Y_s}, \mathbf{X_s}) \in \mathbb{R}^{s*(1+n)}$ is drawn.

2. At each knot, test the global hypothesis of independence between $\mathbf{Y_s}$ and $\mathbf{X_s}$. If the hypothesis cannot be rejected, and independence is assumed, the growth of the tree is stopped in the respective branch. If the hypothesis is rejected, in accordance with a predefined confidence level, the association of each dependent variable with the independent variable is tested and the variable with the highest association, as measured by the highest statistical significance (p value), is chosen as the variable to split on.

3. On the variable with the highest association, the point for the best binary split is chosen as the value of the variable to split on which maximizes the test statistics for association. The data in the respective knot is split by that value as in the classic Random Forests.

4. The steps are repeated within each tree for all trees in the forest until the global null hypothesis can not longer be rejected or another stopping criteria, alike a minimum number of observations in the respective knots, applies.

The statistical tests in this framework do rely on the distribution of the underlying data, however, this distribution is not known. Moreover it would be most inconvenient to have it as a prerequisite for the application of the method. The problem can be solved by the application of permutation testing, where all possible permutations of the values in the learning sample are tested. For more details on the permutation tests, please refer to the framework of conditional inference and permutation tests, developed by (Strasser and Weber 1999) and implemented in the R based cforest algorithm by Hothorn et al. (2006a). Conclusively it is not required to know or estimate the distributions of the variables or risk factors.

For the binary class problem researched in this paper, $Y \in \{0, 1\}$ with class 0 (profit) and 1 (loss), the following test-statistics are applied:

1. First step is the general linear statistic to measure the association between $\mathbf{Y}$ and $\mathbf{X_j}$:

   $$\mathbf{T}_j(\mathcal{O}_s, \mathbf{w}) := vec\left(\sum_{i=1}^{n} w_i g_j(X_{ij}) e_2(Y_i)\right)$$

   The variable to split on is the X_{j*} with $j* = argmin_{j=1,\ldots,m} P_j$ with $P_j = P_{H_0^j}(c(T_j(\mathcal{O}_s, \mathbf{w}), \mu_j, \Sigma_j) \leq c(t_j, \mu_j, \Sigma_j) | S(\mathcal{O}_s, \mathbf{w}))$ with $c_{quad}(\mathbf{T}_j, \mu_j, \Sigma_j) = (\mathbf{T}_j - \mu_j)\Sigma_j^+(\mathbf{T}_j - \mu_j)^T$ Σ is the covariance matrix, Σ^+ is the Moore-Penrose inverse of the covariance matrix, while μ is the mean and S is the permutation of the responses as developed by (Strasser and Weber 1999). Due to the application of these statistics on permutations of the samples, the statistics are conditioned on them.
   In the case of classification the function g_j is the identity mapping or the zero vector with the value one at the level k if a nominal variable with K levels is used ($e_K(k)$). The vec-operator turns a matrix by column-wise combination into a column vector.

2. If the aggregated P value of each T_j test for association cannot be rejected, thus if basically no P value is lower than a predefined level the classification tree is stopped. Hothorn et al. (2006a) suggest to use Bonferroni adjusted P-values or min-P-values.

3. Once the variable, X_{j*}, with the highest association to the dependent variable is found, a similar test statistic is applied: Find best split value on the chosen variable by maximizing the test statistic over all possible subsets of the set of values:

   $$A^* = max_A c_{quad}(\mathbf{T}^A_{j^*}, \mu^A_{j^*}, \Sigma j^{*A})$$

   with

   $$\mathbf{T}^A_{j*}(\mathcal{O}_s, \mathbf{w}) := vec\left(\sum_{i=1}^{n} w_i I(X_{j*i} \in A) e_2(Y_i)\right)$$

The conditional inference framework above is shown by Strobl et al. (2007) to be un-biased in the choice of splitting-variables. The chosen variables can thus be interpreted by the application of the permutation variable importance measure: This importance measure assesses the difference between the prediction of correctly forecast events before and after a respective variable has been randomly permuted.

For further details on the theory of the algorithm of the conditional inference framework, please refer to Strasser and Weber (1999) and Hothorn et al. (2006b).

2.2 Proximity Measures

The focus in this paper is on the proximity measurement, which is a measurement specifying a concept of distance between two observations in a Random Forests model. The idea is simple: For each pair of observations, $\mathcal{O}_i$ and $\mathcal{O}_j$, their proximity ρ_{ij} is defined as the average number of final knots in all trees in the forest where both observations are in the same final knot.

Definition 1. *For two observations $\mathcal{O}_i$ and $\mathcal{O}_j$, $i, j \in \{1, .., m\}$, in a Random Forest with K trees, the proximity ρ_{ij} is defined as*

$$\rho_{ij} := \frac{1}{K}\sum_{k=1}^{K}\sum_{\tilde{k}\in\tilde{K}_k} I(\mathcal{O}_i \in \tilde{k})I(\mathcal{O}_j \in \tilde{k}). \tag{1}$$

$\tilde{K}_k$ is the set of final knots of the k-th tree in the forest and $\tilde{k} := \{\mathcal{O}_{1\tilde{k}}, ..., \mathcal{O}_{m\tilde{k}}, i\tilde{k} \in \{1, ..., m\}\}$ is a single final knot within this set. Thus $\tilde{k}$ is the set of observations which end up in this specific final knot after running through the classification tree. I is the indicator function.

The following characteristic applies:

$$\rho_{jj} = 1 \;\; \forall j \in \{1, ..., M\}, \qquad 0 < \rho_{ij} < 1 \;\; \forall i \neq j \in \{1, ..., M\} \tag{2}$$

For further and more detailed analysis, the overall proximity measure ρ, needs to be broken down into proximities of events and non-events only: the conditional downside and upside proximities.

Definition 2. *Again $Y \in \{0, 1\}$, with the classes 0 (profit) and 1 (loss), then for a chosen Y:*

$$K_i^Y := \sum_{k=1}^{T} I(T_k(\mathcal{O}_i) = Y|Y) \tag{3}$$

K_i^Y is the number of final knots of class Y, in which observation $\mathcal{O}_i$ has ended up in a modeled forest. As before, T is the total number of trees in the respective Random Forests.
The conditional downside proximity of observations $\mathcal{O}_i$ and $\mathcal{O}_j$ is the frequency of joint occurrence of the observations in the same event knots given that observation $\mathcal{O}_i$ is in an event knot:

$$\underline{\rho}_{ij} := P(Y_j = 1|Y_i = 1, \mathcal{O}_i) = \frac{1}{K_i^1}\sum_{k}^{T} I(T_k(\mathcal{O}_i) = 1)I(T_k(\mathcal{O}_j) = 1)I(\tilde{k}_k(\mathcal{O}_i) = \tilde{k}_k(\mathcal{O}_j)) \tag{4}$$

while T_k is the kth tree in a forest and $\tilde{k}_k$ is the function that allocates an end knot to an observation put through a specific tree T_k:

$$\tilde{k}_k(\mathcal{O}_i) \in \tilde{K}_k = \{\tilde{k1}_k, ..., \tilde{k}_{kl}\} \tag{5}$$

with $\tilde{K}_k$ being the set of end knots in tree classifier T_k.

The counterpart to the conditional downside proximity is the conditional upside proximity considering final non-event knots only. For observations $\mathcal{O}_i$ and $\mathcal{O}_j$ it is defined as

$$\overline{\rho}_{ij} = P(Y_j = 0|Y_i = 0, \mathcal{O}_i) = \frac{1}{K_i^0}\sum_{k}^{T} I(T_k(\mathcal{O}_i) = 0)I(T_k(\mathcal{O}_j) = 0)I(\tilde{k}_k(\mathcal{O}_i) = \tilde{k}_k(\mathcal{O}_j)) \tag{6}$$

$\overline{\rho}$ and $\underline{\rho}$ are the m*m matrices of the conditional proximities. The elements of the diagonals of the matrices are 1 while the remaining entries describe how often two observations are in the same event- or non-event knots with regard to the amount of the respective knots of one of the two observations. Consequently, the conditional proximity matrices are not symmetrical.

Additionally, the hedging effect, ρ^h, two observations can have with regard to each other can be modeled by proximities: The hedging effect of an observation with regard to another observation is higher the more likely its performance is to move in the opposite direction within the same environment. The hedging effect will add accuracy to the optimization of portfolio returns.

Definition 3.

$$\rho_{ij}^h = \frac{K_{ij}^{0-1}}{K} \tag{7}$$

with

$$K_{ij}^{0-1} = \sum_k^K I(T_k(\mathcal{O}_i) = 0) I(T_k(\mathcal{O}_j) = 1) \tag{8}$$

which is the amount of final knots per tree which are of the opposite class for two specific observations.

2.3 Derivation of the Proximity based Optimization Measure

In this paper a portfolio is defined as a vector of weights ω_i for m assets, $i \in 1, ..., m$. Each weight equally represents the investment in the respective asset. The total available amount of investments is 1. Thus the sum of the weights is 1 at all times, $\sum_{i=1}^m \omega_i = 1$.

Also, any Random Forests model is fitted to the returns of such a portfolio using a set of risk indicators as independent variables. From this Random Forests model the defined proximities are drawn establishing a direct link between the proximities, the assets, the performance and the risk drivers.

An interdependent risk return ratio of an asset portfolio is usually measured as the ratio of the expected revenue and the risk contribution of the assets to the portfolio. The classic Sharpe ratio (SR) is traditionally defined as the sum of the weighted expected spread, E(S), divided by the square root of the summed and weighted covariance of all assets (Sharpe 1994):

$$SR := \frac{\omega E(S)}{\sqrt{\omega^T cov \omega}} \tag{9}$$

The weights ω are chosen to maximize the Sharpe ratio. Since the risk in the form of the covariance is in the denominator, it follows that the higher the covariance is in comparison to the expected spread, the lower the weight of the respective asset should be. The Sharpe ratio is based on an equilibrium framework, where investors rationally maximize ratio between return and risk. Assuming all necessary assumptions to hold, the Sharpe ratio would thus outperform any other risk return measure.

However, as a wide literature on the Sharpe ratio has already explored, the assumptions do not hold and alternative measures trying to capture the non normal nature of assets returns, can in fact dominate the Sharpe ratio. In the following a risk return ratio is developed that moves away from the standard deviation or similar concepts as risk measure and follows a utility based approach focusing on bond portfolios held for regulatory purposes. As such the maximization of returns contingent on the levels of the portfolios risk drivers are the center of attention of an investor: an investor would thus maximize the ratio of his expected positive return (r^+) and his expected negative return (r^-) given a set of risk indicators $\mathcal{O}$, $\frac{E(r^+|\mathcal{O})}{E(r^-|\mathcal{O})}$. This ratio can be translated into a Random Forests proximity based framework in which a Random Forests model is fitted to the performance of the assets in the portfolio or such which have prospect to enter the portfolio given a set of risk indicators. Then, given this set of risk indicators, positive and negative returns can be incurred by a joint performance of the assets, conditional on the risk indicators (assets are in the same final know of a tree) or independently of the risk indicators (assets are in disjunctive final knots of a tree, but the knots are of the same class). This can be expressed by proximities derived from the Random Forests model on the portfolios return:

$$\frac{E(r^+)\overline{\rho} + E(r^+)\overline{\rho}'}{E(r^-)\underline{\rho} + E(r^-)\underline{\rho}'} \tag{10}$$

where $\overline{\rho}'$ describes the joint probability of assets to exhibit a positive return but being in two disjunctive knots of a tree ($\underline{\rho}'$ respectively for negative returns). However such a ratio does not explicitly consider diversification. Left without constraints at all, an optimization procedure could suggest that the portfolio is best concentrated in one asset. Diversification can be measured by the hedging effect, ρ^h. Considering diversification to be a desirable characteristic of a portfolio, the investor wants to maximize the positive return of diversification effects in relation to negative returns, resulting in the following formula:

$$\frac{E(r^+)\overline{\rho} + E(r^+)\overline{\rho}' + E(r^+)\rho^h}{E(r^-)\underline{\rho} + E(r^-)\underline{\rho}'} \tag{11}$$

reordering the formula yields the following ratio:

$$\frac{E(r^+)}{E(r^-)}\left(\frac{\overline{\rho} + \rho^h}{\underline{\rho}}\frac{\overline{\rho}'}{\underline{\rho}'}\right) \tag{12}$$

For simplifications two changes to the formula are suggested, first instead of the ratio of expected positive returns to expected negative returns it is more customary to use the expected spread E(S). Second the ratio of the joint probabilities of positive and negative return in different knots is set to a constant. Note that this is possible without direct consequence to the $\overline{\rho}$ since $1 - \overline{\rho} = \underline{\rho} + \rho^h + \overline{\rho}' + \underline{\rho}'$. To further simplify the approach the constant is set to 0. Thus the following proximity based risk return ratio results:

$$\text{Proximity based Risk Return Ratio} := E(S)\left(\frac{\overline{\boldsymbol{\rho}} + \boldsymbol{\rho^h}}{\underline{\boldsymbol{\rho}}}\right) = \frac{E(S)}{\frac{(\underline{\boldsymbol{\rho}})}{(\overline{\boldsymbol{\rho}}) + \boldsymbol{\rho^h}}} \tag{13}$$

with the proximity based risk ratio being:

$$\text{Proximity based Risk Ratio} := \boldsymbol{\rho}^{risk} = \frac{(\underline{\boldsymbol{\rho}})}{(\overline{\boldsymbol{\rho}}) + \boldsymbol{\rho^h}} \tag{14}$$

The proximity variants $\underline{\rho}$, $\overline{\rho}$, and $\rho^{\mathbf{h}}$ are m*m matrices, as is the proximity based risk return ratio itself.

Maximizing such a risk return ratio by weighting the observations would clearly reduce portfolio risk.

However, the downside proximity measure is closely related to actual losses, meaning that if no losses are observed, the measure might have a tendency to be low or even zero. On the other hand, in such cases the conditional upside proximity must be larger than zero. Thus the ratio of the downside proximities divided by the upside proximity must be larger than zero. This ratio is the risk of an asset's value to move, in terms of performance, down with the portfolio relative to its chance to move up. The larger this proximity based risk return ratio, the more an asset contributes to a portfolio loss and less to a portfolio gain.

The applied standard optimization formula is then the multiplication of the ratio matrix by the weight vectors:

$$argmin_{\mathbf{X}=(X_1,\ldots,X_N)}(\mathbf{X}^T\frac{(1+\underline{\boldsymbol{\rho}})}{(1+\overline{\rho})+\rho^{\mathbf{h}}}\mathbf{X})=\sum_{i=1}^{N}X_i\sum_{j=1}^{N}\boldsymbol{\rho}_{ij}^{risk}X_j \tag{15}$$

When optimizing a portfolio, the proximity based risk ratio optimizes the interaction of the assets in the following way: for a given asset A1, the ratio shows whether an asset A2 tends to exhibit a loss under equal circumstances in relation to its tendency to exhibit a profit under equal circumstances plus its probability to be profitable while asset A1 exhibits a loss. If this risk ratio is low, the risk and return ratio is higher (assuming a fixed, expected positive return) and thus the weight assigned to asset A2 will increase. If out of N assets, given asset A1, most will perform well, then the weight of asset A1 will also increase.

2.4 Optimization Algorithm

The portfolios are optimized by maximizing the chosen risk return ratio with the method of general nonlinear optimization using the augmented lagrange multiplier described by Yinyu Ye (Ye 1989). This method works for inequality and equality constrained nonlinear programming by minimizing problems with large numbers of variables. The method has inner and out iterations, similarly to sequential or convex quadratic programming.

For more details on the method, please refer to (Ye 1989).

2.5 Benchmark Ratios

A total of 9 additional optimization ratios are used to benchmark the performance of the proximity based risk return ratio. The benchmarks are mostly taken from Cogneau and Huebner's (Cogneau and Hubner 2009) collection. Within their work, Cogneau and Huebner have defined four general types of portfolio optimization measures: ratios of performance and risk, incremental return measures, preference based measures and measures of market timing. This paper considers eight performance and risk ratios as well as one preference based measure: The Sharpe ratio, Adjusted Sharpe ratio, Modified Sharpe ratio, Sortino ratio, Sortino-Sartchell ratio, Upside Potential ratio, Mean Absolute Deviation, Omega Sharpe ratio and the Prospect ratio. For details the reader is referred to Cogneau and Hubner (2009). As the estimation of expected returns is often criticized for its inaccuracy, the performance of the proximity based ratio without return is compared to the success of the minimum variance approach (DeMiguel et al. 2009b).
Thus, apart from the Sharpe ratio which was discussed above, the following benchmarks are taken into account. Because most of these are well known, there will be only a brief description of each. The descriptions follow Cogneau and Hubner (Cogneau and Hubner 2009) closely:

2.5.1 Risk and Performance Ratios

Variants of the Sharpe Ratio
The **Adjusted Sharpe Ratio** was introduced by (Pezier and White 2008). It adjusts for skewness and kurtosis by incorporating a penalty factor for negative skewness and excess kurtosis - both factors which are commonly observed in portfolios.
The formula of the Adjusted Sharpe ratio (ASR) is as stated below (Pezier and White 2008):

$$ASR_i := SR_i(1 + \frac{Skewness}{6}SR_i - \frac{ExcessKurtosis}{24}SR_i^2) \tag{16}$$

A negative skewness and a positive kurtosis thus reduce the value of the ASR in comparison to the SR. However, if the returns are really normally distributed, the skewness and kurtosis are 0 per definition and the ASR is equal to the SR as expected.

Ratios of half and semi variance
Within this category, the most widely used measure is the **Sortino Ratio** (Sortino and VanderMeer 1991) because of its flexibility. It subtracts a reserve return in the numerator and considers the same reserve return in the computation of the semi-variance in the denominator.

A refined variation is the Sortino-Satchell ratio (Sortino and Satchell 2001), in which the semi-variance related to a reserve return is replaced by lower partial moment of order q. It coincides with the Sortino Ratio when q = 2. The introduction of a power index permits the consideration of the investors degree of risk aversion: in practice, a value of q = 0.8 is used to describe an aggressive investor and q = 2.5 for a conservative investor.

Ratios based on VaR and CVaR
The **Modified Sharpe ratio** (Favre and Galeano 2002) uses a modified Value at Risk as denominator instead of the volatility. The modified Value at Risk likewise takes skewness and kurtosis into account by a Cramer-Fisher expansion of the standard Value at Risk quantile $F^{-1}(\alpha)$ (F being the cumulative distribution function of the respective distribution):

$$\begin{aligned} F^{-1}(\alpha)_{CramerFisher} = F^{-1}(\alpha) + \frac{1}{6}(F^{-1}(\alpha)^2 - 1)skewness + \frac{1}{24}(F^{-1}(\alpha)3 \\ - 3F^{-1}(\alpha))Kurtosis - \frac{1}{36}(2F^{-1}(\alpha)^3 - 5F^{-1}(\alpha))skewness^2 \end{aligned} \tag{17}$$

Ratios of gain and shortfall aversion
The **upside potential ratio** proposed by Sortino et al. (Sortino et al. 1999) relies on the idea that the numerator is the expected return above the reserve return and can be seen as the potential of success. The denominator is the downside risk as calculated in the Sortino Ratio. Unlike the Sortino Ratio, the UPR uses the same reference rate for evaluating both profits and losses. Furthermore, the UPR increases with its numerator, which measures the expected return above minimum acceptable return and decreases as its denominator, the downside risk, increases. The UPR therefore delivers performance outputs that conform to the wishes of the investors obtaining rise potential while protecting against losses.

Other absolute risk ratios
A possibility is to consider the mean absolute deviation in the denominator, as in the **mean absolute deviation** ratio of Konno and Yamazaki (Konno and Yamazaki 1991). This ratio is more robust to outliers than the Sharpe ratio.

Bernardo and Ledoit [2000] introduce a measure defined as the ratio of the expectation of the positive part of the returns divided by the expectation of the negative part. The Bernardo-Ledoit gain-loss ratio was rebranded to Omega by Shadwick and Keating (Keating and Shadwick 2002). The ratio can be interpreted as the quotient of a call option and a put option, both having an exercise price equal to the reserve return. The omega is replacing the covariance in the Sharpe ratio leading to the **omega Sharpe Ratio**.

2.5.2 Preference based ratios

For the **Prospect Ratio** (Bacon 2008) value is assigned to gains and losses rather than to final assets. The probabilities are also replaced by decision weights which are generally lower than probabilities.

Any potentially proposed benchmark assets or risk free rate (flat), as inputs to the performance ratios, are set to 0.

3 Application to Portfolio Optimization

To assess the performance of the proximity based risk return ratio as a portfolio optimization measure, it is compared to its benchmarks on three types of analysis. First the performance is measured on three standard and stress scenarios to assess the capabilities of the model under different circumstances. Second, the assets paths are simulated by 1,000,000 Monte Carlo iterations for a time window of 10 years. The optimization is applied to these simulated paths that lead to results which are more focused on the stability of the model. Third, the proximity based risk return ratio and its benchmarks are applied to an empirical bond portfolio. The last test focuses on empirical representativeness using real world data to model the return of government bonds. Although the proposed measure is aiming at optimizing government bond portfolios the steady decline of interest rates during the period under scrutiny renders the capital gain component of the bonds consistently positive and heuristic based optimization methods tend to dominate equilibrium based methods which could bias the results of an empirical study. This paper applied two approaches to remedy a potential bias, first, as can be seen in the appendix, the used annual government bond yields during the selected time window do not follow a trend nor is the yield consistently positive, second the analysis on simulated asset returns is designed to replicate a portfolio of assets which promise higher returns for higher risk taking without any detectable heuristic trend.

The following table 1 summarizes the three tests in comparison.

To evaluate the performance of the chosen optimization strategies in comparison to each other, several measures are computed for all analyzes:

1. The average in-sample and out-of-sample performance of each strategy.

2. The in-sample and out-of-sample portfolio Sharpe ratio (PF Sharpe and Oos PF Sharpe).

Description of Performance Tests

Type of Analysis	Stability	Scenarios	Representative
Scenario Analysis	One year of data is simulated. Random variations are present	Covers performance, correlation, skew and kurtosis scenarios	Lacks empirical representation, is artificially created but balances the bond focus of the other analyzes
Monte Carlo Analysis	The results are stable due to the simulation of 1'000'000 paths	Scenarios are empirically defined	Simulation based on empirical data. Bond focused
Empirical Analysis	Fewer data points due to lack of empirical data	Scenarios are empirically defined	No assumptions are introduced, only empirical data is used. Bond focused

Table 1: Description of the three applied types of performance test.

3. The average turnover rate, defined as the average changes in individual daily asset weights.

For simplicity, the portfolio optimization is performed on the returns of the assets only, without considering the total value of the resulting portfolio or changes in the later. Thus, the turnover rate for the one/N strategy, the strategy with stable and equal weights for each asset, is always zero. It needs to be noted that for simplicity, transaction costs are also omitted. For this reason it has to be taken into consideration that large differences in the turnover rate could lead to different optimization results if transaction costs were included. This depends on the size of the transaction costs of course.

For the scenario analysis it is additionally tested whether the portfolio Sharpe ratios (resulting from the application of the strategies in- and out-of-sample) are statistically distinguishable. The return to VaR ratio on a portfolio level is also considered for the simulation and the empirical analysis. The resulting portfolios for the simulation and empirical analysis are done annually for roughly 10 years. Although this is the time window this paper is focused on, there are not enough data points to conduct statistical testing in this two cases.

As the proximities in the ratio can theoretically be zero in the case of certain, very 'distant' assets, the value 1 is added in the implementation to both proximities, which is not changing the ratio between compared assets but prevents infinite values which are hard to cope with.

3.1 First Application: Scenario Analysis

Scenario Design

The model is first tested against its benchmarks on a portfolio of simulated asset returns. The simulated portfolio consists of three assets. Three is the minimum amount of assets for a meaningful application since for only two assets the weights tend to converge to 0.5 each. This is a consequence of equation 15. For each of the three assets the daily log returns for each trading day over five years are simulated. The underlying distributions are characterized by scenarios on the first four central moments and the correlation of the asset returns: three scenarios are tested focusing on no correlation, positive and negative correlations as well as low and high skew and kurtosis.

To receive meaningful, consistent results, a Random Forests has to be constructed and therefore a simple correlated simulation will not work. Thus, the following is done: two beta distributed support-variables with t data points are simulated and normalized, the results are used to construct the returns of asset 1, asset 2 and asset 3. To control the correlation between the three asset return vectors the value of the asset 1 returns is added to the asset 2 returns, or the asset 2 returns to the asset 3 returns, times a correlation factor. The beta distribution facilitates a low skew and high kurtosis within the resulting distributions:

AssetReturn(i) := RandomBetaNumber(N, Alpha 1, Beta 1) + RandomBetaNumber(N, Alpha 2, Beta 2) - AdjustmentFactor(i)*(mean(RandomBetaNumber(t, Alpha 1, Beta 1))+ mean(RandomBetaNumber(N, Alpha 2, Beta 2))) + (AssetReturn(i-1) * CorrelationFactor)

Additionally factors are calibrate to keep the returns mostly on the positive side (AdjustmentFactor := {0.99, 0.98, 0.99}). With this construction the support variables are systematically connected with the asset returns. Binding the three asset return variables to a single dataset and defining class 1 as the cases where the delta from one observation to the next (return) over all assets is negative and class 0 otherwise, a simple Random Forests model can be run from which the indirectly simulated and required proximity measures are drawn. Varying the input: mean, standard deviation and correlation factor, the asset return distributions in table 2 are produced. Considering the relationship between the returns and

Descriptive Statistics of the Scenarios and Assets

Scenario	Asset1 Mean	Std	Skew	Kurt.	Asset2 Mean	Std	Skew	Kurt.	Asset3 Mean	Std	Skew	Kurt.	Corr
Base Case	0.02	0.15	-0.73	1.15	0.04	0.16	-1.08	1.55	0.05	0.16	-1.34	2.03	0
+Correlation	0.02	0.15	-0.63	0.68	0.05	0.19	-0.76	1.27	0.08	0.22	-0.96	1.26	0.51
-Correlation	0.02	0.14	-0.48	0.56	0.03	0.17	-0.86	1.55	0.04	0.19	-0.99	1.36	-0.21

Table 2: Summary of the mean, skew and kurtosis off all three assets in all three scenarios including the average correlation.

the standard deviation between assets within the different scenarios, it can be seen that with only one exception the riskier assets promise higher returns. This is necessary to simulate a relationship on which measures with a variance based risk concept can sensibly optimize a portfolio. On the other hand the nested application of random numbers assures that a Random Forests model build on the simulated data cannot perfectly reproduce the results: The Random Forests model drawn in the simulation analysis has an average classification error of around 15% to 30%.

An optimization is run on the three vectors of simulated asset returns for all three scenarios to find the weights of the assets in a portfolio to maximize the proximity based risk return ratio and its benchmarks. The simulation is repeated to produce 1275 observations, representing data for 5 years while the used proximities and correlations at point t are always built using the simulated information of the number of observations representing the preceding 3 months. The portfolio optimization is also constrained by a maximum of a 45% weight for each of the three assets to encourage stability.

Scenario Analysis Results

For all strategies, it is tested whether the Sharpe ratio of the respectively optimized portfolio is significantly different from the Sharpe ratio of the portfolio received by optimizing the proximity based risk return ratio. Because the distributions are chosen to have negative skew and very high kurtosis, standard statistical tests are not meaningfully applicable (for example, the Jobson and Korkie test with a correction by Memmel (Ledoit and Wolf 2008)). The chosen statistical test is the two sided studentized circular block bootstrap test developed by Ledoit and Wolf (2008) for non-normal distributions. The implementation is done using the R-code provided by Ledoit and Wolf [1]. If the proximity based risk return ratio outperforms an alternative measure and is significantly different it can be assumed that its dominance is not due to random deviations and it performs actually better.

Table 3 summarizes the results of the asset return scenario with no correlation and correlations close to zero. The proximity based risk return ratio exhibits the highest in-sample returns as well as

Results of Scenario application: Base Case

Base Case -Skew +Kurtosis	Average Return	Oos Return	Stdev	PF Sharpe	Oos Stdev	Oos PF Sharpe	Turnover Rate	p	Oos p
Proximity based Ratio	0.039	0.0356	0.101	0.3862	0.103	0.346	0.1437		
Standard Share Ratio	0.0337	0.0309	0.0976	0.3451	0.0997	0.3096	0.5329	0.004	0.0198
Modified Sharpe Ratio	0.0347	0.0311	0.0963	0.3601	0.0992	0.3136	0.5238	0.0553	0.0395
Adjusted Sharpe Ratio	0.0326	0.0342	0.1037	0.3144	0.1011	0.3381	0.6681	0.004	0.6324
Sortino Satchell Ratio	0.0346	0.0311	0.0965	0.3584	0.0993	0.3127	0.5342	0.0514	0.0395
MAD	0.0326	0.0344	0.1029	0.3172	0.1015	0.3392	0.6544	0.004	0.7589
Omega Sharpe Ratio	0.0351	0.0309	0.1006	0.349	0.1034	0.2985	0.6729	0.0119	0.004
Prospect Ratio	0.0348	0.0316	0.0998	0.3491	0.1029	0.3074	0.6661	0.0119	0.0277
Sortino Ratio	0.0348	0.0308	0.0979	0.356	0.1012	0.3042	0.6109	0.0277	0.0079
Upside Potential Ratio	0.0344	0.0313	0.0953	0.3608	0.098	0.3193	0.4828	0.0395	0.1028
One/N	0.033	0.033	0.0957	0.3448	0.0957	0.3448	0	0.004	0.2609

Table 3: The table contains the results of the application of all performance ratios to the base case scenario. The return is in shown as share of the portfolio value.

out-of-sample returns. Its application also leads to the highest in sample portfolio Sharpe ratios while out-of-sample the One/N strategy, MAD and Adjusted Sharpe ratio exhibit a similar portfolio Sharpe ratio. In-sample, the proximity based risk return ratio is statistically distinguishable from almost all other measures on a confidence level of 5% with exception of the Modified Sharpe ratio and the Sortino Satchell ratio where the p values are 5.5% and 5.1% respectively. It is thus dominates all competitors with a significance level of 10%. However, out-of-sample the proximity based risk return ratio is not significantly distinguishable in four of the cases. On the other hand the portfolio which is optimized by the proximity based risk return ratio is restructured less during its lifetime than the other portfolios. Transaction costs will thus have a lower impact on the relative performance of the proximity based risk return ratio in comparison to its competitors.

In the case of positively correlated asset returns the proximity based risk return ratio outperforms the alternative strategies likewise with regard to portfolio return and portfolio Sharpe ratio, in- and out-of-sample. The turnover rate is also clearly lowest in table 4. The in-sample as well as out-of-sample portfolio results of the proximity based risk return ratio are statistically distinguishable on a confidence level of 1% with the exception of the out-of-sample Omega Sharpe ratio results (where the p value is 0.016).

[1]The code can be optioned by http://www.ledoit.net/jef2008_abstract.htm.

Results of Scenario application: Positive Correlation

Positive Correlation -Skew +Kurtosis	Average Return	Oos Return	Stdev	PF Sharpe	Oos Stdev	Oos PF Sharpe	Turnover Rate	p	Oos p
Proximity based Ratio	0.0598	0.0575	0.1632	0.3663	0.164	0.3504	0.1008		
Standard Share Ratio	0.0485	0.0483	0.1515	0.3204	0.1498	0.3223	0.4632	0.004	0.004
Modified Sharpe Ratio	0.0488	0.0486	0.1514	0.3225	0.1506	0.323	0.4796	0.004	0.004
Adjusted Sharpe Ratio	0.0476	0.0466	0.154	0.3094	0.1552	0.3002	0.681	0.004	0.004
Sortino Satchell Ratio	0.049	0.0485	0.151	0.3243	0.1505	0.3223	0.4733	0.004	0.004
MAD	0.0472	0.0469	0.1539	0.3068	0.1555	0.3016	0.6769	0.004	0.004
Omega Sharpe Ratio	0.0491	0.0493	0.1554	0.3158	0.1527	0.323	0.655	0.004	0.0158
Prospect Ratio	0.0491	0.0485	0.1537	0.3191	0.1517	0.3197	0.6361	0.004	0.004
Sortino Ratio	0.0491	0.0488	0.1519	0.323	0.1506	0.3242	0.5288	0.004	0.0079
Upside Potential Ratio	0.0491	0.0483	0.1511	0.3248	0.15	0.3221	0.4551	0.004	0.004
One/N	0.0478	0.0478	0.1491	0.3204	0.1491	0.3204	0	0.004	0.004

Table 4: The table contains the results of the application of all performance ratios to the positive correlation scenario. The return is in shown as share of the portfolio value.

In case of negative correlation (table 5) the proximity based risk return ratio again exhibits higher levels on the in- and out-of-sample portfolio returns and portfolio Sharpe ratios with exception of the out-of-sample One/N strategy.

Results of Scenario application: Negative Correlation

Negative Correlation -Skew +Kurtosis	Average Return	Oos Return	Stdev	PF Sharpe	Oos Stdev	Oos PF Sharpe	Turnover Rate	p	Oos p
Proximity based Ratio	0.0343	0.0292	0.0874	0.3924	0.0886	0.3298	0.1041		
Standard Share Ratio	0.0286	0.0247	0.0801	0.357	0.0831	0.2966	0.4855	0.004	0.0079
Modified Sharpe Ratio	0.0288	0.0244	0.0813	0.3548	0.0835	0.2918	0.5211	0.0988	0.1621
Adjusted Sharpe Ratio	0.0233	0.0277	0.0932	0.2502	0.0893	0.3099	0.6719	0.004	0.3399
Sortino Satchell Ratio	0.0282	0.0246	0.0805	0.3504	0.0826	0.2976	0.5028	0.0711	0.0119
MAD	0.0247	0.0267	0.0924	0.2676	0.0885	0.3019	0.6442	0.004	0.2451
Omega Sharpe Ratio	0.0263	0.0255	0.0897	0.2927	0.091	0.2799	0.6757	0.004	0.0949
Prospect Ratio	0.0257	0.0255	0.0902	0.2849	0.0912	0.2798	0.6766	0.004	0.0751
Sortino Ratio	0.0277	0.0246	0.0829	0.3343	0.0851	0.2886	0.5722	0.0237	0.253
Upside Potential Ratio	0.0283	0.0243	0.0793	0.3565	0.0816	0.298	0.4661	0.1462	0.1976
One/N	0.0266	0.0266	0.0744	0.3583	0.0744	0.3583	0	0.0514	0.1976

Table 5: The table contains the results of the application of all performance ratios to the negative correlation scenario. The return is in shown as share of the portfolio value.

The in-sample performance of the proximity based ratio is mostly statistically significantly better than the performance of its competitors on a confidence level of 5%, with exception of the Modified Sharpe, the Sortino Satchell and the Upside Potential ratio as well as the One/N strategy. However, the out-of-sample performance of the proposed model is only statistically significantly better than the performance of the Sharpe ratio and the Sortino Satchell ratio on a 5% confidence level.

The result that the proximity based performance ratio outperforms its competitors could be based on a hypothetical tendency of the suggested ratio to simply apply daily weights of 45% for the two best performing assets and 10% for the least performing asset. In other words, the ratio could simply follow the estimated expected return. However, ignoring all risks and optimizing the portfolio based on the expected return only, the strategy would lead to an overall portfolio return (in-sample) of 7.9% (base case), 8.82% (positive correlated case) and 8.09% (negative correlated case) which are clearly different from the received results (3.9%, 5.9%, 3.43% respectively) . Overall the proximity based risk return ratio has higher returns which are often statistically significant and the turnover rate is always considerably lower. However, caution should be exercised in portfolios with considerable negative correlation. It is important to assure the accuracy of the Random Forests model as the results above indicate that the proximities perform less well.

Additionally, as table 6 shows, the proximity based risk ratio performs equally well in all three scenarios as the minimum variance strategy as the resulting PF Sharpe ratios are the highest in three cases for each ratio. This indicates that the proximity based risk ratio contributes to the results of the proposed ratio, as shown in table 6.

Comparison of Risk Ratios

Scenario	Average Return	Oos Return	Stdev	PF Sharpe	Oos Stdev	Oos PF Sharpe
Base Case	0	0	0	0	0	0
Proximity based Ratio	0.033	0.0342	0.0955	0.3453	0.0948	0.361
Minimum Variance	0.0328	0.0327	0.09	0.3649	0.0908	0.3606
Positive Correlation	0	0	0	0	0	0
Proximity based Ratio	0.0479	0.0493	0.161	0.2976	0.1617	0.3048
Minimum Variance	0.0409	0.0407	0.1508	0.2709	0.1515	0.2689
Negative Correlation	0	0	0	0	0	0
Proximity based Ratio	0.0276	0.0266	0.0784	0.352	0.0806	0.3294
Minimum Variance	0.0249	0.0246	0.0688	0.3619	0.0696	0.3538

Table 6: The table describes the results of the comparison of the proximity based risk ratio vs the minimum variance approach. The return is in shown as share of the portfolio value.

3.2 The Monte Carlo and Empirical Analysis

To set the analysis on more representative grounds empirical data is used. The empirical data is the basis for two approaches, both using the same proximity related measures, calculated on the empirical dataset. First, because the empirical data offers relatively few data points, a Monte Carlo simulation is run. The parameters of the simulation of the correlated asset paths are derived from the empirical portfolio. And second, as empirical benchmark, the actually incurred PnL and observed risk indicators are taken into account. The empirical benchmark has relatively few data points but embedded in the scenario analysis and the Monte Carlo simulation exhibits additional information.

The Dataset

To construct a suitable empirical dataset that allows the sensible running of a Random Forests model, sufficient independent variables are required. Thus two datasets were used: The dependent variable is the end of day price of 10 year government bonds issued by the respective country. The 20 years of data in the sample encompasses information from Australia, Canada, France, Germany, Ireland, Japan, the Netherlands, New Zealand, Sweden, the United Kingdom and the United States.

For the independent variables, the public and online available data of the World Bank, 'World Development Indicators & Global Development Finance' is used [2].

The independent indicators are selected from currently applied theories on GDP growth, such as tax raising, public spending, monetary policy, the liberty of the economic environment, the workforce and its education and international trade. Indicators with more than 33% missing values are excluded. Indicators that cannot be easily compared between countries, such as indicators measured in local currency or other absolute values, are also not included. In numbers, 104 indicators are chosen between 1990 and 2011[3]. The large amount of indicators in the model can easily be coped with by Random Forests and as Biau (2012) shows, there will be no distortion from variables with no predictive power as long as some others do have. The indicators and their descriptions are listed in appendix **??** to the document. Since the Random Forests based recursive conditional partitioning does not overfit (Breiman and Cutler 2005), many more indicators could theoretically be included.

The recursive conditional partitioning framework is used as a classification algorithm, thus the dependent variable has to be binary. An event, which is a loss in the respective bond price, is defined as whenever the respective bond price is lower than in the previous year. The dependent variable Y will take the value 1 for an adverse price movement and 0 for other instances.

To comply with risk management imposed investment limits within financial institutions each asset has a maximum weight of 25%. Again, short-selling is not allowed.

Parameters of the Random Forests-cforest Model

The cforest algorithm implemented in the R 'party' package is applied, using the following parameter settings: quadratic test-statistics with a splitting criterion of a variable which is associated with at least 99% significance, a minimum sum of weights in the knot of twice the weight of non-event cases and a minimum sum of weights of each of the subsequent knots of the weight of event cases. This choice is expected to lead to high accuracy while no empty knots are possible. The weights themselves are the inverse proportion of the amount of events or non-events in the dataset. The number of sampled variables tried at each split is set to the square root of the number of independent variables. The choices with regard to sampled variables and weights are as suggested by Breiman and Cutler (2005). For the stability of the results, 5,000 trees are run for each forest.

Model Calibration

The Random Forests-cf model is calibrated to minimize the classification error of the fitted forest by specific calibration of the class weights. As mentioned, class weights are the inverse proportion of the amount of events or non-events in the dataset (Breiman and Cutler 2005).
The analysis is done on a historic rolling window of 10 years for the Monte Carlo simulation and the empirical analysis. The bootstrap sample in the forests is set to be 63.2%, as suggested by Strobl et al. (2007).

[2] Online in internet: http://data.worldbank.org/indicator
[3] With an average of 9% missing values between 1999-2011.

3.3 Second Application: Monte Carlo Analysis

The simulation is done in accordance with the Credit Metrics approach of Gupton et al. (1997) and Saunders and Allen (2002): the parameters are the mean returns and standard deviation of the empirical dataset as well as the correlation matrix which is built from the dataset. The migration matrix is extracted from publicly available S&P data. In total, 1,000,000 asset paths are simulated. The paths describe the valuation profit and losses due to migration transition or default. Yet, since an event in the analysis is not the default of an asset, the default probabilities in the migration matrix are replaced by an estimation of the empirical probability to incur a loss for each rating[4].

The S&P migration matrix is scaled as to assure that the sums of the rows still add up to 1 after insertion of the alternative probability of default estimates.

On the 10 year rolling window of data, the portfolio optimized by the proximity based ratio performs best in- as well as out of the sample with regard to the ratio of return to value at risk. In all other cases the Omega Sharpe ratio performs equally well or slightly better (table 7).

Results of Monte Carlo Simulation Analysis

Performance Comparison for 10 Year Time Window	Realized Average Return	Realized VaR to Return Ratio	Realized Portfolio Sharpe Ratio	Oos Average Return	Oos VaR to Return Ratio	Oos Portfolio Sharpe Ratio	Turnover Rate
Proximity Based Ratio	0.31	1.36	0.61	0.33	1.64	0.64	0.39
Standard Sharpe Ratio	0.19	0.55	0.39	0.23	0.69	0.49	0.42
Modified Sharpe Ratio	0.19	0.36	0.36	0.22	0.57	0.43	0.35
Adjusted Sharpe Ratio	0.24	0.66	0.52	0.25	0.72	0.56	0.14
Sortino Satchell Ratio	0.23	0.75	0.45	0.2	0.65	0.41	0.31
MAD	0.21	0.39	0.38	0.29	0.85	0.55	0.61
Omega Sharpe Ratio	0.33	1.2	0.63	0.33	1.15	0.69	0.68
Prospect Ratio	0.21	0.59	0.42	0.25	0.78	0.52	0.35
Sortino Ratio	0.22	0.62	0.43	0.24	0.67	0.49	0.37
Upside Potential Ratio	0.29	0.97	0.53	0.33	1.17	0.63	0.31
One/N	0.28	0.2	0.3	0.28	0.2	0.3	0

Table 7: The table presents the results of the Monte Carlo simulated optimized portfolio performance with a 10 year rolling window. The return is in shown as share of the portfolio value.

The turnover rate is comparatively low, however, it is clearly lower than the turnover rate of the portfolio structured by the Omega Sharpe ratio.

3.4 Third Application: Empirical Analysis

In-sample of the empirical data, the portfolio using the proximity based ratio has the highest return but also has very high risks leading to the return to VaR and portfolio Sharpe ratio being at the end of the ranking. In the out-of-sample analysis, however, the proximity based ratio exhibits the second highest return and the highest risk return ratios (table 8).

[4]The estimates are defined as the per rating average of the empirical probabilities per asset of a negative daily return deviation which is larger than the standard deviation of all daily returns. This method assures that higher probabilities of a loss are associated with lower ratings.

Results of the Empirical Analysis

Scenario	Realized Average Return	Realized VaR to Return Ratio	Realized Portfolio Sharpe Ratio	Oos Average Return	Oos VaR to Return Ratio	Oos Portfolio Sharpe Ratio	Turnover Rate
Proximity Based Ratio	0.37	1.19	0.6	0.27	0.57	0.4	0.34
Standard Sharpe Ratio	0.35	2.01	0.81	0.16	0.27	0.29	0.39
Modified Sharpe Ratio	0.35	1.18	0.63	0.17	0.29	0.26	0.45
Adjusted Sharpe Ratio	0.32	1.52	0.72	0.15	0.26	0.27	0.57
Sortino Satchell Ratio	0.33	1.88	0.74	0.16	0.25	0.27	0.38
MAD	0.19	0.21	0.22	0.28	0.47	0.38	0.44
Omega Sharpe Ratio	0.36	2.3	0.79	0.15	0.26	0.28	0.41
Prospect Ratio	0.34	2.19	0.78	0.15	0.27	0.28	0.4
Sortino Ratio	0.34	1.92	0.76	0.16	0.28	0.29	0.39
Upside Potential Ratio	0.31	1.57	0.63	0.13	0.22	0.24	0.5
One/N	0.28	0.2	0.3	0.28	0.2	0.3	0

Table 8: The table presents the results of the portfolio optimization based on empirical data only. The return is in shown as share of the portfolio value.

The turnover rate of the proximity based strategy is the smallest of all alternatives, apart from the One/N strategy which has no turnover.

3.5 Discussion of Monte Carlo and Empirical Analysis

Both, the analysis on the empirical data as well as the simulation on the empirical data are based on too few data points to conduct statistical testing for significant differences within the portfolio Sharpe ratios. However, considering the performance ranking, the proximity based ratio is on average most often within the top two measures and, together with the Omega Sharpe ratio, it is most often the best performing ratio. The proximity based ratio performs especially well in the out-of-sample case. Also the proximity based ratio has empirically outperformed the one/N strategy, in- as well as out-of-sample.

4 Application to Portfolio Management

4.1 Risk Contribution

A useful measure of portfolio risk management is the risk contribution of individual assets to the portfolio risk. Risk contribution is usually defined as the covariance between an asset and the whole portfolio divided by the unexpected loss of the portfolio. However, the exact definition depends on the measure used for portfolio optimization as the risk contribution should match the way it is implicitly already measured within the optimization approach. This leads to an alternative risk contribution measure defined by the proximity based ratio.

Definition 4. *Classic Risk Contribution Measure of Asset i:*

$$RC_i = \frac{cov(AssetReturn_i, PortfolioReturn)}{VaR_{Portfolio}} \tag{18}$$

Proximity Based Risk Contribution Measure of Asset i:

$$RC(\rho)_i = \sum_{j}^{N} \frac{(1+\underline{\rho}_{ij})}{(1+\overline{\rho}_{ij}) + \rho_{ij}^{h}} \tag{19}$$

4.2 Risk Indicator Importance

As pointed out in the introduction the conditional recursive partitioning forest is mainly applied to allow drawing reliable importance scores from it to assess the relative contribution of individual risk indicators. Thus when fitting the Random Forests model to the bond prices, importance measures for each indicator were also calculated.

Strobl et al. (2009) defines the applied variable importance measure in the following way: *Importance is defined by randomly permuting the values of a predictor variable and thus breaking its original association with the response. Thus, a reasonable measure for variable importance is the difference in prediction accuracy before and after permuting a variable, averaged over all trees.*

As per definition, the lowest importance value would be zero, meaning that the variable is simply not adding any information to the classification. Obviously the difference can be negative if a noise variable suddenly adds to the prediction after permutation although it did not do so before. Strobl et al. (2009) state that the range from the lowest negative importance measure to its absolute positive value constitutes the threshold for important variables as non important variables deviate randomly around 0. This is considered as a rule of thumb. However, due to the theoretical truncation of the distribution of the permuted importance values at zero, a positive bias within the measure of the importance of variables without any association to the dependent variable seems plausible. Thus, with a symmetric rule of thumb too many variables are considered to contribute significantly while in fact they do not.

To test this, a suitable number of randomly generated variables and a randomly generated bi-variate response variable were generated. The predictor variable is not associated with the other variables by design, meaning any association is random. A first analysis was comprised of a cforest and consecutive importance measure construction using 250 observations, 5,000 trees and 20 to 2,000 variables. The graphical interpretation of the above results confirms the negatively skewed shape of the distribution, showing a fat tail on the right. As the distribution of the values in the importance measure depend on the variable as well as the model, it seems to be a conservative approach to focus on the highest percentile of importance scores of the variables for an analysis.

Thus, for each year in the analysis, the distribution of the importance measures was built and the 90% percentile was taken. The top three risk indicators which were in the 90% percentile in most of the analyzed years are identified for all years. Their occurrence in the 90% percentile, on the whole sample and during the financial crisis only, is summarized in table 9. As the importance measure basically measures the contribution of an indicator to classify an observation as event or non-event, the most important variables in a specific year clearly give some indication of why the bond prices went down in that year. Table 9 shows that indicators of the countries banks liquid reserves are important for the performance of the bond prices. In times when the whole portfolio incurred a loss, indicators related

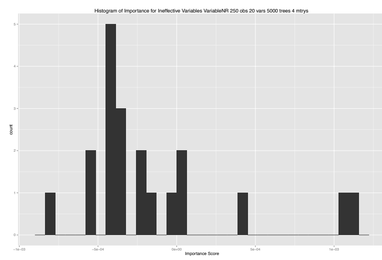

Figure 1: Histogram of Importance for Ineffective Variables VariableNR 250 obs 20 vars 5000 trees 4 mtrys

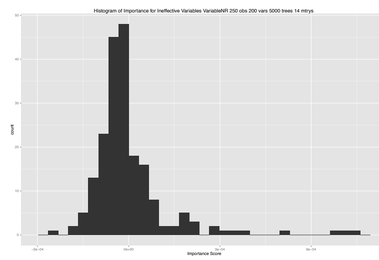

Figure 2: Histogram of Importance for Ineffective Variables VariableNR 250 obs 200 vars 5000 trees 14 mtrys

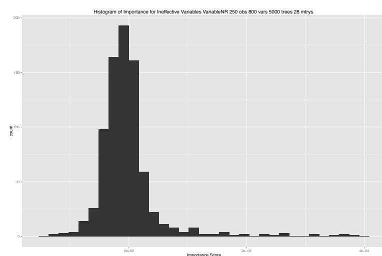

Figure 3: Histogram of Importance for Ineffective Variables VariableNR 250 obs 800 vars 5000 trees 28 mtrys

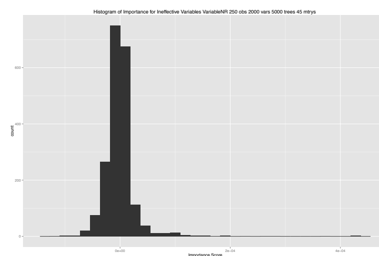

Figure 4: Histogram of Importance for Ineffective Variables VariableNR 250 obs 2000 vars 5000 trees 45 mtrys

Results of Importance Analysis

	Frequency of importance above 90% percentile	
Three most important variables	In the whole sample	During financial crisis
Bank liquid reserves to bank assets ratio (%)	100%	100%
Claims on other sectors of the domestic economy (% of GDP)	95%	100%
Interest payments (% of expense)	91%	87.5%

Table 9: The table describes how often the 3 most important risk indicators' importance scores are within the 90% percentile of the distribution of importance scores within different time frames.

to claims on other sectors of the domestic economy[5] gained importance, while the interest payments on the bonds were more important in calm times. The important risk drivers are an additional means to manage the risk in the portfolio with regard to its single assets. An asset that has a high value (or low value, depending on the direction of the effects of the indicator in the model) on an important indicator should be considered more risky irrespective of its risk contribution.

[5]This indicator includes gross credit from the financial system to households, non-profit institutions serving households, non-financial corporations, state and local governments and social security funds.

4.3 Proximity Based Risk Management Table

The identified risk indicators with a high importance measure can be used to enhance risk management driven decisions with regard to the portfolio composition. The most important indicators are taken to demonstrate this. With each of these risk indicators a recursive conditional partitioning tree with only one split is built. In each of this trees the approximate value which most homogeneously splits the observations in events and non events is searched for. Values of the respective variables above this thresholds are expected to be more often associated with rising bond prices while values below are associated with losses. This results in a risk driven portfolio management overview. This overview is summarized for 1992 as an example:

Table 10: Example of Portfolio Management Table

CountryName	Date	Performance	RC	Estimated Risk Return Ratio	Bank liquid reserves to bank assets ratio (%)	Claims on other sectors of the domestic economy (% of GDP)	Interest payments (% of expense)
Australia	1992	0.41	6.15	0.0095			
Canada	1992	0.99	5.91	0.0098			
France	1992	0.51	6.52	0.0061			
Germany	1992	0.98	5.65	0.007			
Ireland	1992	-0.48	6.77	0.0086	FLAG	FLAG	
Japan	1992	1.27	5.85	0.0068			
Netherlands	1992	0.98	5.92	0.0067			
NewZealand	1992	0.59	5.87	0.0099			
Sweden	1992	-0.98	7.12	0.0082	FLAG		FLAG
UnitedKingdom	1992	0.59	5.78	0.0069			
UnitedStates	1992	0.25	6.02	0.0066			

The complete tables used for analysis in this paper are in the appendix. Overall, the Random Forests model cannot attribute a single value to a specific risk indicator that distinguishes events from non events as it is a complex model of various trees and many more knots where splits are performed. The importance measurements are just pointing out which indicator has the most influence in the algorithm. None the less, the risk indicators can be used as a warning flag for enhanced risk. The performance of the flags is very good around the bond price crisis of 1994, good on the whole time window of 1991-2011 and less good but still better than random decision making during the financial crisis from 2007 onwards. The following table shows the ratio of the correctly predicted gains to all gains, the ratio of the correctly predicted losses to all losses and the ratio of correctly predicted losses to all observations for each time period and indicator, which are lower than the respective threshold. The last column, FLAG Hits, describes the accuracy of the indicator flag itself.

If any two flags are risen, 57% of the observations estimated as losses actually are losses. If all three flags rise, the accuracy rises to 73%.

Thus, after the portfolio composition is set by the weights from the optimization, each asset in the portfolio, or potential asset that could be included in the portfolio, can be managed by the calculated risk contribution measure and risk to return ratio as well as the risk flags on the important indicators.

Table 11: Average Accuracy of Risk Indicators

		Correctly predicted gains %	Correctly predicted loss %	FLAG Hits
1991-2011	Bank liquid reserves to bank assets ratio (%)	0.75	0.56	0.53
	Claims on other sectors of the domestic economy (% of GDP)	0.84	0.52	0.62
	Interest payments (% of expense)	0.73	0.57	0.52
1992-1995	Bank liquid reserves to bank assets ratio (%)	0.97	1	0.93
	Claims on other sectors of the domestic economy (% of GDP)	0.87	0.92	0.75
	Interest payments (% of expense)	0.81	1	0.68
2007-2009	Bank liquid reserves to bank assets ratio (%)	0.61	0.4	0.46
	Claims on other sectors of the domestic economy (% of GDP)	0.72	0.53	0.62
	Interest payments (% of expense)	0.22	0.67	0.42

5 Conclusion

The aim of this paper is to introduce a portfolio optimization measure focusing on the characteristics of portfolios held primarily for regulatory reasons:

- The Portfolio is comprised of mostly high quality assets; a majority of which are government bonds.
- No short selling within the portfolio is allowed and a maximum/minimum weight limit for the assets is applied.
- The portfolio is not acquired for trading purposes, thus a rather stable inventory with a risk management focus is preferred.
- Usually smaller financial institutions do not have a rich data warehouse to model a portfolio management approach.

The introduced portfolio performance measure tailored to this characteristics is built on a Random Forests analysis focusing particularly on the proximities and variations of these.

The approach allows the modeling of the interdependence of assets in a portfolio as an alternative to covariance and correlations. Proximities can be modeled with data on different assets from one year only, provided events and non-events are present in the dataset and thus do not require a long internal data history. To model the proximities, large amounts of external and publicly available data can be used to tackle the problem of data limitations, which is often encountered in especially smaller financial institutions. Since Random Forests does not overfit, this characteristic allows the application of the proximity based performance ratio with big data. Moreover, applying Random Forests is independent of the underlying distributions of profits and losses as well as of all risk factors.

The analysis has shown that the proximity based performance ratio is not influenced by the skewness and kurtosis of the portfolio returns.

Considering its performance against benchmarks, the proximity based ratio has been shown to outperform other models. In most cases it exhibits higher profit and lower risk than its benchmarks, the difference is often statistically significant. The proximity based performance ratio is better suited to optimize the considered portfolios rather than the tested peers especially considering the turnover rate which is considerably lower than for the peers in all cases also the ones where the proximity based risk return ratio is not statistically distinguishable.

The fitting of the Random Forests model to the portfolio returns using large numbers of risk indicators allows directly linking the proximities with risk indicators and the portfolios assets. Thus the approach adds portfolio risk management instruments by allowing for the identification of the most important drivers of the portfolio risk and it introduces a risk contribution measure for each asset in a portfolio.

As such it is especially suitable for smaller institutions often lacking the database for suitable mathematical models as it promotes the extensive usage of external data, a direct link to prudent risk management and low transaction costs.

References

Bacon, C., 2008: *Practical portfolio performance measurement and attribution.* John Wiley & Sons, Inc.

Biau, G., 2012: Analysis of a random forests model. *Journal of Machine Learning Research*, **13**, 1063–1095.

Bonner, C., 2014: Preferential regulatory treatment and banks demand for government bonds. *De Nederlandsche Bank Working Paper No. 433, Online in internet: http://dx.doi.org/10.2139/ssrn.2474058.*

Breiman, L. and A. Cutler, 2005: Random forests. *Available online at: http://www.stat.berkeley.edu/users/breiman/ RandomForests/cc_home.htm (accessed 25.07.2005).*

Brodie, J., I. Daubechies, C. D. Mol, D. Giannone, and I. Loris, 2008: Sparse and stable markovitz portfolios. *European Central Bank Working Paper Series, Available online at: http://ssrn.com/abstract_id=1258442 (accessed 09.04.2015)*, **936**.

Cherny, A. and D. Madan, 2007: New measures for performance evaluation. *Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=955472 (accessed 09.04.2015).*

Cogneau, P. and G. Hubner, 2009: The 101 ways to measure portfolio performance. *Available at SSRN: http://ssrn.com/abstract=1326076 or http://dx.doi.org/10.2139/ssrn.1326076.*

DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal, 2009a: A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, **55, 5**, 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal, 2009b: Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, **22, 5**.

Fastrich, B., S. Paterlini, and P. Winker, 2014: Constructing optimal sparse portfolios using regularization methods. *Working Paper, Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2169062. (accessed 09.04.2015).*

Favre, L. and J.-A. Galeano, 2002: Mean-modified value-at-risk optimization with hedge funds. *Journal of Alternative Investment*, **5**.

Gupton, G., C. Finger, and M. Bhatia, 1997: Creditmetrics: The benchmark for understanding credit risk. *Technical Document, New York, NY: JP Morgan Inc.*

Hothorn, T., K. Hornik, and A. Zeileis, 2006a: party: A laboratory for recursive partytioning. *Available online at: http://cran.r-project.org/web/packages/party/vignettes/party.pdf (08.01.2015).*

Hothorn, T., K. Hornik, and A. Zeileis, 2006b: Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, **15, 3**, 651–674.

Iscoe, I. and A. Kreinin, 2000: Default boundary problem. *Technical report ARPS, Algorithmics Inc., Toronto*, **00, 03**.

Iscoe, I., A. Kreinin, and D. Rosen, 1999: An integrated market and credit risk portfolio model. *Algo Research Quarterly*, **2, 3**, 2137.

Jothimani, D., R. Shankar, and S. Yadav, 2014: A big data analytical framework for portfolio optimization. *http://epic.is.cityu.edu.hk/wibf/docs/WIBF14-Papers/wibf14_submission_14.pdf.*

Keating, C. and W. F. Shadwick, 2002: A universal performance measure. *Journal of Performance Measurement*, **6, 3**, 59–84.

Konno, H. and H. Yamazaki, 1991: Mean-absolute deviation portfolio optimization model and its application to tokyo stock market. *Management Science*, **37, 5**, 519–531.

Ledoit, O. and M. Wolf, 2008: Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, **15**, 850–859.

Pezier, J. and A. White, 2008: The relative merits of investable hedge fund indices and of funds of hedge funds in optimal passive portfolios. *The Journal of Alternative Investments*, **10 (4)**, 37–49.

Saunders, A. and L. Allen, 2002: Credit risk measurement: New approaches to value at risk and other paradigms. *New York: John Wiley and Sons.*

Sharpe, W. F., 1994: The sharpe ratio. *Journal of Portfolio Management*, **Fall**, 49–58.

Sortino, F. A. and S. E. Satchell, 2001: *Managing downside risk in financial markets.* Batterworth-Heinemann Finance, Oxford.

Sortino, F. R., V. D. Meer, and A. Plantinga, 1999: The dutch triangle. *Journal of Portfolio Management*, **26, 5**, 50–58.

Sortino, F. R. and R. VanderMeer, 1991: Downside risk. *Journal of Portfolio Management*, **17, 4**, 27–31.

Strasser, H. and C. Weber, 1999: On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics. Available online at: http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub. wu-wien.ac.at:epub-wu-01_94c.*, **8**, 220–250.

Strobl, C., A. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, **8**, 1471–2105.

Strobl, C., J. Malley, and G. Tutz, 2009: An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol Methods*, **14(4)**, 323348.

Tasche, D., 2008: Capital allocation to business units and sub-portfolios: the euler principle. *Available online at: http://arxiv.org/pdf/0708.2542.pdf (accessed 09.04.2015).*

Ye, Y., 1989: Solnp users' guide - a nonlinear optimization program in matlab. *Department of Management Sciences, College of Business Administration, University of Iowa.*

Curve Discussion

The applied annual government bond returns, defined as the delta in NPV per year, do not exhibit a specific trend within the duration of the sample:

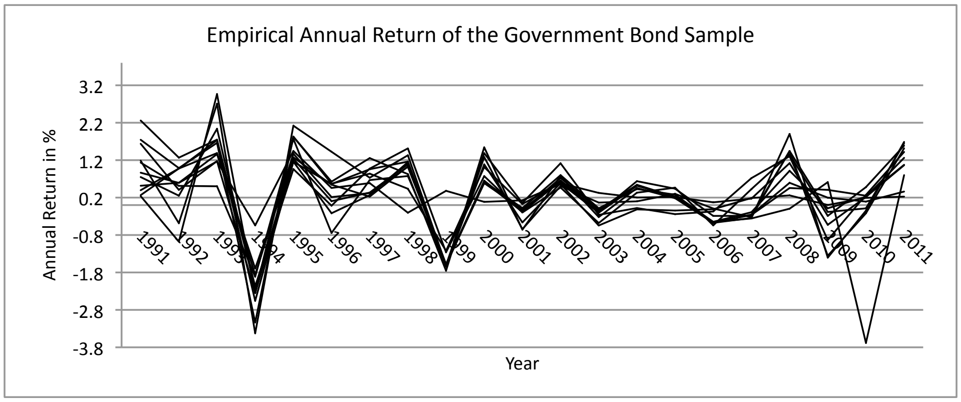

Figure 5: Graphical analysis of potential trends in the sampled bond returns

The included 10 year government bonds are from Australia, Canada, France, Germany, Ireland, Japan, the Netherlands, New Zealand, Sweden, the United Kingdom and the United States.

Risk Management Tables

Table 12: Example of Portfolio Management Table

CountryName	Date	Performance	RC	Estimated Risk Return Ratio	Bank liquid reserves to bank assets ratio (%)	Claims on other sectors of the domestic economy (% of GDP)	Interest payments (% of expense)
Australia	1992	0.41	6.15	0.0095			
Canada	1992	0.99	5.91	0.0098			
France	1992	0.51	6.52	0.0061			
Germany	1992	0.98	5.65	0.007			
Ireland	1992	-0.48	6.77	0.0086	FLAG	FLAG	
Japan	1992	1.27	5.85	0.0068			
Netherlands	1992	0.98	5.92	0.0067			
NewZealand	1992	0.59	5.87	0.0099			
Sweden	1992	-0.98	7.12	0.0082	FLAG		FLAG
UnitedKingdom	1992	0.59	5.78	0.0069			
UnitedStates	1992	0.25	6.02	0.0066			
Australia	1993	2.04	6.16	0.0094			
Canada	1993	1.17	6.58	0.0088			
France	1993	1.66	5.96	0.0067			
Germany	1993	0.5	6.86	0.0058			
Ireland	1993	1.74	5.98	0.0097			
Japan	1993	2.71	8.76	0.0045			
Netherlands	1993	1.75	6.81	0.0058			
NewZealand	1993	1.39	6.06	0.0096			
Sweden	1993	1.17	6.49	0.009			
UnitedKingdom	1993	2.96	9.99	0.004			
UnitedStates	1993	1.36	5.45	0.0073			
Australia	1994	-1.69	10.67	0.0055	FLAG	FLAG	FLAG
Canada	1994	-2.25	10.66	0.0055	FLAG		
France	1994	-2.35	8.93	0.0045	FLAG	FLAG	FLAG
Germany	1994	-3.14	10.73	0.0037	FLAG	FLAG	FLAG
Ireland	1994	-1.91	10.85	0.0054	FLAG	FLAG	FLAG
Japan	1994	-2.17	10.74	0.0037	FLAG	FLAG	FLAG
Netherlands	1994	-3.43	10.78	0.0037	FLAG	FLAG	FLAG
NewZealand	1994	-2.56	10.77	0.0054	FLAG	FLAG	FLAG
Sweden	1994	-1.86	10.54	0.0055	FLAG	FLAG	FLAG
UnitedKingdom	1994	-0.54	10.43	0.0038	FLAG	FLAG	FLAG
UnitedStates	1994	-2.35	10.7	0.0037	FLAG	FLAG	
Australia	1995	1.16	10.26	0.0057			
Canada	1995	1.24	10.39	0.0056			
France	1995	1.28	10.54	0.0038			FLAG
Germany	1995	1.83	8.98	0.0044			FLAG
Ireland	1995	1.82	10.39	0.0056			
Japan	1995	1.77	9.51	0.0042			
Netherlands	1995	0.96	10.58	0.0038			
NewZealand	1995	2.12	10.35	0.0056			
Sweden	1995	1.16	10.56	0.0055			
UnitedKingdom	1995	1.37	10.22	0.0039			
UnitedStates	1995	1.45	9.97	0.004			
Australia	1996	0.46	5.72	0.0102			
Canada	1996	0.56	5.81	0.01			
France	1996	0.1	5.76	0.0069		FLAG	FLAG
Germany	1996	0.55	5.89	0.0068		FLAG	FLAG
Ireland	1996	0.57	5.72	0.0102			
Japan	1996	0.63	5.7	0.007			
Netherlands	1996	-0.75	7.86	0.0051			
NewZealand	1996	-0.02	7.95	0.0073	FLAG	FLAG	FLAG
Sweden	1996	1.42	5.75	0.0101			
UnitedKingdom	1996	-0.22	7.11	0.0056			FLAG
UnitedStates	1996	0.21	5.67	0.007		FLAG	FLAG

List of applied independent risk indicators

Table 13: Example of Portfolio Management Table

CountryName	Date	Performance	RC	Estimated Risk Return Ratio	Bank liquid reserves to bank assets ratio (%)	Claims on other sectors of the domestic economy (% of GDP)	Interest payments (% of expense)
Australia	1997	0.32	6.03	0.0097			
Canada	1997	0.59	5.93	0.0098			
France	1997	0.96	6.1	0.0065		FLAG	FLAG
Germany	1997	0.33	6.03	0.0066		FLAG	FLAG
Ireland	1997	0.23	6.16	0.0094			
Japan	1997	0.84	6.08	0.0065			
Netherlands	1997	1.25	6.16	0.0065			
NewZealand	1997	0.67	9.12	0.0064			
Sweden	1997	0.97	8.71	0.0067			
UnitedKingdom	1997	0.73	6.15	0.0065		FLAG	
UnitedStates	1997	0.27	8.92	0.0045			
Australia	1998	1.15	5.5	0.0106			
Canada	1998	1.02	5.78	0.0101			
France	1998	-0.2	8.5	0.0047			
Germany	1998	1.16	5.73	0.0069			
Ireland	1998	1.09	5.8	0.0069			
Japan	1998	1.03	6.07	0.0096			
Netherlands	1998	0.44	5.8	0.0069			
NewZealand	1998	0.83	5.67	0.0103			
Sweden	1998	0.78	6.32	0.0092			
UnitedKingdom	1998	1.51	5.73	0.0069			
UnitedStates	1998	1.33	5.68	0.007			
Australia	1999	-1.6	9.22	0.0063	FLAG	FLAG	FLAG
Canada	1999	-1.64	9.05	0.0064	FLAG		
France	1999	-1.59	9.02	0.0044	FLAG	FLAG	FLAG
Germany	1999	0.38	6.16	0.0065	FLAG	FLAG	FLAG
Ireland	1999	-1.6	9.03	0.0044	FLAG	FLAG	FLAG
Japan	1999	-1.58	8.37	0.007			
Netherlands	1999	-1.59	9.06	0.0044	FLAG	FLAG	FLAG
NewZealand	1999	-1.22	8	0.0073	FLAG	FLAG	FLAG
Sweden	1999	-1.75	8.77	0.0066	FLAG	FLAG	
UnitedKingdom	1999	-1.61	8.79	0.0045	FLAG	FLAG	FLAG
UnitedStates	1999	-1.26	9.57	0.0042	FLAG		FLAG
Australia	2000	0.77	10.06	0.0058			
Canada	2000	1.09	9.98	0.0058			
France	2000	1.28	9.81	0.0041		FLAG	
Germany	2000	0.65	9.66	0.0041		FLAG	
Ireland	2000	0.09	7.17	0.0055		FLAG	
Japan	2000	0.63	9.58	0.0061			
Netherlands	2000	0.62	9.23	0.0043	FLAG	FLAG	
NewZealand	2000	0.59	9.66	0.006			
Sweden	2000	1.01	9.74	0.006		FLAG	
UnitedKingdom	2000	1.54	10.02	0.004			
UnitedStates	2000	1.39	9.09	0.0044			
Australia	2001	0.06	5.58	0.0104	FLAG	FLAG	FLAG
Canada	2001	-0.2	8.55	0.0068	FLAG		
France	2001	-0.46	8.91	0.0045	FLAG	FLAG	FLAG
Germany	2001	-0.64	9.12	0.0044		FLAG	FLAG
Ireland	2001	-0.16	8.92	0.0045	FLAG	FLAG	FLAG
Japan	2001	0.13	5.57	0.0104			
Netherlands	2001	-0.08	8.6	0.0046	FLAG	FLAG	FLAG
NewZealand	2001	-0.18	8.33	0.007	FLAG	FLAG	FLAG
Sweden	2001	-0.12	9.56	0.0061	FLAG	FLAG	FLAG
UnitedKingdom	2001	0.03	5.91	0.0067	FLAG		FLAG
UnitedStates	2001	-0.66	8.81	0.0045	FLAG		

Table 14: Example of Portfolio Management Table

CountryName	Date	Performance	RC	Estimated Risk Return Ratio	Bank liquid reserves to bank assets ratio (%)	Claims on other sectors of the domestic economy (% of GDP)	Interest payments (% of expense)
Australia	2002	0.78	8.55	0.0068		FLAG	FLAG
Canada	2002	1.12	6.21	0.0094			
France	2002	0.47	8.54	0.0047		FLAG	FLAG
Germany	2002	0.59	8.5	0.0047		FLAG	FLAG
Ireland	2002	0.51	8.87	0.0045		FLAG	FLAG
Japan	2002	0.72	8.55	0.0068			
Netherlands	2002	0.56	6.17	0.0064		FLAG	FLAG
NewZealand	2002	0.8	8.67	0.0067			FLAG
Sweden	2002	0.64	7.97	0.0073		FLAG	
UnitedKingdom	2002	0.66	8.43	0.0047			FLAG
UnitedStates	2002	0.58	6.47	0.0061			
Australia	2003	0.32	5.31	0.011			FLAG
Canada	2003	-0.32	7.79	0.0075			
France	2003	-0.26	7.36	0.0054			FLAG
Germany	2003	-0.48	8.27	0.0048			FLAG
Ireland	2003	-0.2	8.33	0.0048			FLAG
Japan	2003	0.07	5.71	0.0102			FLAG
Netherlands	2003	-0.11	8.67	0.0046			FLAG
NewZealand	2003	-0.55	7.73	0.0075			FLAG
Sweden	2003	-0.09	8.01	0.0073			FLAG
UnitedKingdom	2003	-0.2	8.53	0.0047			FLAG
UnitedStates	2003	-0.16	8.45	0.0047			FLAG
Australia	2004	0.54	8.34	0.007		FLAG	
Canada	2004	0.2	6.48	0.0061	FLAG		
France	2004	0.46	7.52	0.0053		FLAG	
Germany	2004	-0.08	8.92	0.0045		FLAG	
Ireland	2004	0.33	8.17	0.0049		FLAG	
Japan	2004	0.64	7.95	0.0073			
Netherlands	2004	0.1	6.11	0.0065		FLAG	
NewZealand	2004	0.54	7.94	0.0073		FLAG	
Sweden	2004	-0.12	9.01	0.0044	FLAG	FLAG	
UnitedKingdom	2004	0.43	7.71	0.0052		FLAG	
UnitedStates	2004	0.52	8	0.005			
Australia	2005	0.22	6.65	0.0088	FLAG	FLAG	FLAG
Canada	2005	0.22	5.85	0.0068	FLAG		
France	2005	0.22	6.56	0.0061	FLAG	FLAG	FLAG
Germany	2005	0.18	5.93	0.0067	FLAG	FLAG	FLAG
Ireland	2005	-0.24	10.1	0.0039	FLAG	FLAG	FLAG
Japan	2005	0.46	6.01	0.0097			
Netherlands	2005	0.44	5.75	0.0069	FLAG	FLAG	FLAG
NewZealand	2005	0.3	5.94	0.0098	FLAG	FLAG	FLAG
Sweden	2005	0.22	6.03	0.0066	FLAG	FLAG	FLAG
UnitedKingdom	2005	-0.15	10.1	0.0039	FLAG	FLAG	FLAG
UnitedStates	2005	0.28	5.83	0.0068	FLAG		
Australia	2006	-0.45	7.4	0.0079		FLAG	
Canada	2006	-0.45	8.17	0.0049	FLAG		
France	2006	-0.47	7.77	0.0051		FLAG	
Germany	2006	0.07	5.43	0.0073		FLAG	
Ireland	2006	-0.47	7.76	0.0051		FLAG	
Japan	2006	-0.17	10.39	0.0056			
Netherlands	2006	-0.54	8.24	0.0048		FLAG	
NewZealand	2006	-0.29	7.89	0.0074		FLAG	
Sweden	2006	-0.09	7.72	0.0052	FLAG	FLAG	
UnitedKingdom	2006	-0.48	8.1	0.0049		FLAG	
UnitedStates	2006	-0.11	10.36	0.0038			

Table 15: Example of Portfolio Management Table

CountryName	Date	Performance	RC	Estimated Risk Return Ratio	Bank liquid reserves to bank assets ratio (%)	Claims on other sectors of the domestic economy (% of GDP)	Interest payments (% of expense)
Australia	2007	0.19	6.83	0.0085			FLAG
Canada	2007	-0.36	12.05	0.0033	FLAG		
France	2007	-0.19	12.12	0.0033		FLAG	FLAG
Germany	2007	-0.25	12.21	0.0033		FLAG	FLAG
Ireland	2007	0.17	5.84	0.0068			FLAG
Japan	2007	-0.35	12.1	0.0048			
Netherlands	2007	0.72	6.48	0.0061			FLAG
NewZealand	2007	0.43	7.56	0.0077			FLAG
Sweden	2007	-0.29	12.98	0.0031	FLAG		FLAG
UnitedKingdom	2007	-0.33	12.39	0.0032			FLAG
UnitedStates	2007	-0.26	12.56	0.0032			
Australia	2008	0.46	8.56	0.0068	FLAG		FLAG
Canada	2008	0.26	6.19	0.0064	FLAG		
France	2008	-0.1	8.53	0.0047			FLAG
Germany	2008	0.91	8.11	0.0049			FLAG
Ireland	2008	0.59	8.32	0.0048	FLAG		FLAG
Japan	2008	1.12	6.52	0.0089	FLAG		
Netherlands	2008	1.9	8.24	0.0048	FLAG		FLAG
NewZealand	2008	1.3	6.04	0.0096	FLAG		FLAG
Sweden	2008	1.35	6.56	0.0061			FLAG
UnitedKingdom	2008	1.44	8.79	0.0045	FLAG		FLAG
UnitedStates	2008	1.44	8.18	0.0049			
Australia	2009	-1.35	7.93	0.0073			FLAG
Canada	2009	0.41	5.48	0.0073			
France	2009	0	5.99	0.0066		FLAG	
Germany	2009	0.61	7.28	0.0055		FLAG	
Ireland	2009	-0.09	7.5	0.0078			
Japan	2009	0.2	6.02	0.0097			
Netherlands	2009	-0.53	8.37	0.0048			
NewZealand	2009	-1.4	7.38	0.0079		FLAG	FLAG
Sweden	2009	-0.93	8.23	0.0048		FLAG	FLAG
UnitedKingdom	2009	-0.29	8.18	0.0049		FLAG	
UnitedStates	2009	-0.18	8.24	0.0048			
Australia	2010	-0.09	9.14	0.0064	FLAG	FLAG	
Canada	2010	-0.22	9.5	0.0042			
France	2010	0.25	6.01	0.0066	FLAG	FLAG	
Germany	2010	0.19	6.16	0.0065		FLAG	
Ireland	2010	-3.68	6.73	0.0147	FLAG		
Japan	2010	0.22	7.36	0.0079			
Netherlands	2010	0.09	6.05	0.0066			
NewZealand	2010	0.22	7.79	0.0075	FLAG	FLAG	FLAG
Sweden	2010	-0.15	9.51	0.0042	FLAG	FLAG	FLAG
UnitedKingdom	2010	0.27	7.52	0.0053		FLAG	
UnitedStates	2010	0.48	7.62	0.0052			
Australia	2011	1.59	5.97	0.0067	FLAG		FLAG
Canada	2011	1.65	8.46	0.0047			
France	2011	1.52	8.17	0.0049			
Germany	2011	1.08	6.14	0.0065			
Ireland	2011	0.23	5.86	0.0169	FLAG		
Japan	2011	0.8	7.74	0.0075			
Netherlands	2011	1.26	5.92	0.0067			FLAG
NewZealand	2011	0.36	5.92	0.0098			FLAG
Sweden	2011	1.07	5.86	0.0068	FLAG		FLAG
UnitedKingdom	2011	1.69	7.83	0.0051			
UnitedStates	2011	1.42	5.81	0.0068			

Table 16: Default occurrence estimation accuracy, restricted on most frequent underlying model accuracies

Variable Name	Variable Description	Theory Class	Importance Score in 2010 Model
IC.REG.COST.PC.ZS	Cost of business start-up procedures (% of GNI per capita)	Economic Environment	0.01553
PA.NUS.ATLS	DEC alternative conversion factor (LCU per US$)	Monetary	0.01524
NE.IMP.GNFS.KD.ZG	Imports of goods and services (annual % growth)	International Trade	0.01481
NE.GDI.TOTL.KD.ZG	Gross capital formation (annual % growth)	Spending	0.01295
NE.GDI.TOTL.ZS	Gross capital formation (% of GDP)	Spending	0.01094
NE.EXP.GNFS.KD.ZG	Exports of goods and services (annual % growth)	International Trade	0.01055
FP.CPI.TOTL	Consumer price index (2005 = 100)	Macroeconomic	0.00904
FP.CPI.TOTL.ZG	Inflation, consumer prices (annual %)	Macroeconomic	0.00881
NE.CON.TETC.KD.ZG	Final consumption expenditure, etc. (annual % growth)	Spending	0.00835
NE.GDI.FTOT.ZS	Gross fixed capital formation (% of GDP)	Spending	0.00757
IC.ISV.DURS	Time to resolve insolvency (years)	Economic Environment	0.00633
NY.GDP.DEFL.ZS	GDP deflator (base year varies by country)	Macroeconomic	0.00546
NY.GDP.DEFL.KD.ZG	Inflation, GDP deflator (annual %)	Macroeconomic	0.00479
IC.LGL.PROC	Procedures to enforce a contract (number)	Economic Environment	0.00446
TM.VAL.FOOD.ZS.UN	Food imports (% of merchandise imports)	International Trade	0.00435
TM.VAL.MRCH.XD.WD	Import value index (2000 = 100)	International Trade	0.00398
NE.CON.GOVT.ZS	General government final consumption expenditure (% of GDP)	Spending	0.00378
NY.GNS.ICTR.ZS	Gross savings (% of GDP)	Economic Environment	0.00348
NY.GNS.ICTR.GN.ZS	Gross savings (% of GNI)	Economic Environment	0.00285
FM.LBL.MQMY.GD.ZS	Money and quasi money (M2) as % of GDP	Monetary	0.00282
TX.VAL.MRCH.XD.WD	Export value index (2000 = 100)	International Trade	0.00265
IT.NET.USER.P2	Internet users (per 100 people)	Economic Environment	0.00254
SL.EMP.1524.SP.MA.ZS	Employment to population ratio, ages 15-24, male (%)	Workforce and Education	0.00248
TM.QTY.MRCH.XD.WD	Import volume index (2000 = 100)	International Trade	0.00245
SH.XPD.PCAP.PP.KD	Health expenditure per capita, PPP (constant 2005 international $)	Spending	0.00226
GB.XPD.RSDV.GD.ZS	Research and development expenditure (% of GDP)	Spending	0.00208
IP.JRN.ARTC.SC	Scientific and technical journal articles	Spending	0.00204
FM.LBL.MQMY.ZG	Money and quasi money growth (annual %)	Monetary	0.00199
GC.TAX.TOTL.GD.ZS	Tax revenue (% of GDP)	Tax	0.00183
BX.TRF.PWKR.DT.GD.ZS	Workers' remittances and compensation of employees, received (% of GDP)	Economic Environment	0.00176
GC.XPN.TOTL.GD.ZS	Expense (% of GDP)	Spending	0.00162
SL.EMP.TOTL.SP.MA.ZS	Employment to population ratio, 15+, male (%)	Workforce and Education	0.0016
NE.GDI.TOTL.CD	Gross capital formation (current US$)	Spending	0.00155
NE.DAB.TOTL.ZS	Gross national expenditure (% of GDP)	Spending	0.00144
IT.CEL.SETS.P2	Mobile cellular subscriptions (per 100 people)	Economic Environment	0.0014
NE.RSB.GNFS.ZS	External balance on goods and services (% of GDP)	International Trade	0.00133
NY.GNS.ICTR.CD	Gross savings (current US$)	Economic Environment	0.0012
TX.VAL.TECH.CD	High-technology exports (current US$)	International Trade	0.00111
NY.GDS.TOTL.CD	Gross domestic savings (current US$)	Economic Environment	0.00111
BX.KLT.DINV.CD.WD	Foreign direct investment, net inflows (BoP, current US$)	International Trade	0.0011

Table 17: Default occurrence estimation accuracy, restricted on most frequent underlying model accuracies

Variable Name	Variable Description	Theory Class	Importance Score in 2010 Model
IC.REG.DURS	Time required to start a business (days)	Economic Environment	0.00109
NY.GDS.TOTL.ZS	Gross domestic savings (% of GDP)	Spending	0.00105
IC.LGL.DURS	Time required to enforce a contract (days)	Economic Environment	0.001
NE.GDI.FTOT.CD	Gross fixed capital formation (current US$)	Spending	0.00098
BM.KLT.DINV.GD.ZS	Foreign direct investment, net outflows (% of GDP)	International Trade	0.00094
IC.REG.PROC	Start-up procedures to register a business (number)	Economic Environment	0.00092
IS.AIR.GOOD.MT.K1	Air transport, freight (million ton-km)	Economic Environment	0.00092
TX.VAL.MRCH.WL.CD	Merchandise exports by the reporting economy (current US$)	International Trade	0.0009
SH.XPD.PCAP	Health expenditure per capita (current US$)	Spending	0.0009
GC.XPN.TRFT.ZS	Subsidies and other transfers (% of expense)	Spending	0.00089
SL.EMP.1524.SP.ZS	Employment to population ratio, ages 15-24, total (%)	Workforce and Education	0.00087
SH.XPD.PUBL.ZS	Health expenditure, public (% of GDP)	Spending	0.00087
TM.VAL.MANF.ZS.UN	Manufactures imports (% of merchandise imports)	International Trade	0.00085
NY.TAX.NIND.CD	Net taxes on products (current US$)	Tax	0.00084
TX.VAL.MRCH.CD.WT	Merchandise exports (current US$)	International Trade	0.00084
TM.VAL.MRCH.CD.WT	Merchandise imports (current US$)	International Trade	0.00082
BN.CAB.XOKA.GD.ZS	Current account balance (% of GDP)	International Trade	0.0008
GC.REV.SOCL.ZS	Social contributions (% of revenue)	Tax	0.0008
GC.XPN.COMP.ZS	Compensation of employees (% of expense)	Economic Environment	0.00079
NE.GDI.TOTL.KD	Gross capital formation (constant 2000 US$)	Spending	0.00076
NE.IMP.GNFS.CD	Imports of goods and services (current US$)	International Trade	0.00074
NE.GDI.FTOT.KD	Gross fixed capital formation (constant 2000 US$)	Spending	0.00072
GC.XPN.INTP.ZS	Interest payments (% of expense)	Spending	0.00071
BX.KLT.DINV.WD.GD.ZS	Foreign direct investment, net inflows (% of GDP)	International Trade	0.0007
SH.XPD.PUBL.GX.ZS	Health expenditure, public (% of government expenditure)	Spending	0.0007
BM.GSR.TRAN.ZS	Transport services (% of service imports, BoP)	International Trade	0.00069
NE.EXP.GNFS.CD	Exports of goods and services (current US$)	International Trade	0.00068
NE.CON.PETC.ZS	Household final consumption expenditure, etc. (% of GDP)	Economic Environment	0.00065
SL.EMP.TOTL.SP.ZS	Employment to population ratio, 15+, total (%)	Workforce and Education	0.00065
SL.EMP.1524.SP.FE.ZS	Employment to population ratio, ages 15-24, female (%)	Workforce and Education	0.00065
TX.VAL.MRCH.R5.ZS	Merchandise exports to developing economies in South Asia (% of total merchandise exports)	International Trade	0.00064
NE.RSB.GNFS.CD	External balance on goods and services (current US$)	International Trade	0.00063
SE.PRM.ENRL.FE.ZS	Primary education, pupils (% female)	Workforce and Education	0.00063
TM.VAL.TRAN.ZS.WT	Transport services (% of commercial service imports)	International Trade	0.00063
GC.TAX.OTHR.RV.ZS	Other taxes (% of revenue)	Tax	0.00062
SH.XPD.TOTL.ZS	Health expenditure, total (% of GDP)	Spending	0.00062
BN.GSR.MRCH.CD	Net trade in goods (BoP, current US$)	International Trade	0.00061
TX.QTY.MRCH.XD.WD	Export volume index (2000 = 100)	International Trade	0.00057
TX.VAL.MANF.ZS.UN	Manufactures exports (% of merchandise exports)	International Trade	0.00057
NE.CON.TETC.ZS	Final consumption expenditure, etc. (% of GDP)	Spending	0.00056
SE.PRM.AGES	Primary school starting age (years)	Workforce and Education	0.00055
IP.TMK.TOTL	Trademark applications, total	Economic Environment	0.00054
TT.PRI.MRCH.XD.WD	Net barter terms of trade index (2000 = 100)	International Trade	0.00053
NY.GSR.NFCY.CD	Net income from abroad (current US$)	International Trade	0.00051
SL.TLF.CACT.MA.ZS	Labor participation rate, male (% of male population ages 15+)	Workforce and Education	0.00051
NE.CON.PRVT.PC.KD	Household final consumption expenditure per capita (constant 2000 US$)	Economic Environment	0.00046
FS.AST.DOMS.GD.ZS	Domestic credit provided by banking sector (% of GDP)	Economic Environment	0.00045
ST.INT.TRNR.CD	International tourism, receipts for passenger transport items (current US$)	International Trade	0.00045
TM.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service imports)	International Trade	0.00044
EG.ELC.LOSS.KH	Electric power transmission and distribution losses (kWh)	Economic Environment	0.00043
TX.VAL.MRCH.R3.ZS	Merchandise exports to developing economies in Latin America and the Caribbean (% of total merchandise exports)	International Trade	0.00042
TM.VAL.MMTL.ZS.UN	Ores and metals imports (% of merchandise imports)	International Trade	0.00042
SE.XPD.TOTL.GD.ZS	Public spending on education, total (% of GDP)	Spending	0.0004
TX.VAL.MRCH.R6.ZS	Merchandise exports to developing economies in Sub-Saharan Africa (% of total merchandise exports)	International Trade	0.0004

Table 18: Default occurrence estimation accuracy, restricted on most frequent underlying model accuracies

Variable Name	Variable Description	Theory Class	Importance Score in 2010 Model
TX.VAL.MRCH.RS.ZS	Merchandise exports by the reporting economy, residual (% of total merchandise exports)	International Trade	0.0004
ST.INT.TVLX.CD	International tourism, expenditures for travel items (current US$)	International Trade	0.00038
SE.ENR.PRIM.FM.ZS	Ratio of female to male primary enrollment (%)	Workforce and Education	0.00038
NE.EXP.GNFS.ZS	Exports of goods and services (% of GDP)	International Trade	0.00037
BN.GSR.GNFS.CD	Net trade in goods and services (BoP, current US$)	International Trade	0.00036
ST.INT.XPND.MP.ZS	International tourism, expenditures (% of total imports)	International Trade	0.00036
TX.VAL.MRCH.HI.ZS	Merchandise exports to high-income economies (% of total merchandise exports)	International Trade	0.00036
TX.VAL.FOOD.ZS.UN	Food exports (% of merchandise exports)	International Trade	0.00035
SE.SEC.ENRL.GC.FE.ZS	Secondary education, general pupils (% female)	Workforce and Education	0.00035
TX.VAL.TECH.MF.ZS	High-technology exports (% of manufactured exports)	International Trade	0.00035
NE.CON.GOVT.KD.ZG	General government final consumption expenditure (annual % growth)	Spending	0.00034
BN.KLT.DINV.CD	Foreign direct investment, net (BoP, current US$)	International Trade	0.00034
SE.PRM.ENRR.FE	School enrollment, primary, female (% gross)	Workforce and Education	0.00033
NE.IMP.GNFS.ZS	Imports of goods and services (% of GDP)	International Trade	0.00032
SH.XPD.PRIV.ZS	Health expenditure, private (% of GDP)	Spending	0.00032
BM.GSR.FCTY.CD	Income payments (BoP, current US$)	Economic Environment	0.00032
FS.AST.DOMO.GD.ZS	Claims on other sectors of the domestic economy (% of GDP)	Economic Environment	0.00032
BN.TRF.KOGT.CD	Net capital account (BoP, current US$)	International Trade	0.00031
NY.TRF.NCTR.CD	Net current transfers from abroad (current US$)	International Trade	0.00031
NE.CON.PRVT.CD	Household final consumption expenditure (current US$)	Economic Environment	0.00031
BX.GSR.FCTY.CD	Income receipts (BoP, current US$)	Economic Environment	0.0003
NE.CON.TETC.CD	Final consumption expenditure, etc. (current US$)	Spending	0.0003
BX.GSR.TRVL.ZS	Travel services (% of service exports, BoP)	International Trade	0.00029
BM.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, paid (current US$)	Economic Environment	0.00028
NE.CON.PETC.CD	Household final consumption expenditure, etc. (current US$)	Economic Environment	0.00028
SE.SEC.ENRL.FE.ZS	Secondary education, pupils (% female)	Workforce and Education	0.00028
IT.MLT.MAIN.P2	Telephone lines (per 100 people)	Economic Environment	0.00028
NE.CON.PRVT.PP.KD	Household final consumption expenditure, PPP (constant 2005 international $)	Economic Environment	0.00027
NE.EXP.GNFS.KD	Exports of goods and services (constant 2000 US$)	International Trade	0.00026
SL.TLF.CACT.FE.ZS	Labor participation rate, female (% of female population ages 15+)	Workforce and Education	0.00026
TX.VAL.TRAN.ZS.WT	Transport services (% of commercial service exports)	International Trade	0.00026
ST.INT.XPND.CD	International tourism, expenditures (current US$)	International Trade	0.00025
NE.CON.TOTL.CD	Final consumption expenditure (current US$)	Spending	0.00025
NE.DAB.TOTL.KD	Gross national expenditure (constant 2000 US$)	Spending	0.00025
TM.VAL.INSF.ZS.WT	Insurance and financial services (% of commercial service imports)	International Trade	0.00025
BN.CAB.XOKA.CD	Current account balance (BoP, current US$)	International Trade	0.00025
TX.VAL.MRCH.OR.ZS	Merchandise exports to developing economies outside region (% of total merchandise exports)	International Trade	0.00025
NE.IMP.GNFS.KD	Imports of goods and services (constant 2000 US$)	International Trade	0.00024
FS.AST.CGOV.GD.ZS	Claims on central government, etc. (% GDP)	Economic Environment	0.00024
GC.TAX.YPKG.ZS	Taxes on income, profits and capital gains (% of total taxes)	Tax	0.00024
SE.PRM.ENRL	Primary education, pupils	Workforce and Education	0.00024
IT.CEL.SETS	Mobile cellular subscriptions	Economic Environment	0.00023
TX.VAL.TRVL.ZS.WT	Travel services (% of commercial service exports)	International Trade	0.00023
IS.AIR.DPRT	Air transport, registered carrier departures worldwide	Economic Environment	0.00022
NE.DAB.TOTL.CD	Gross national expenditure (current US$)	Spending	0.00022
ST.INT.TRNX.CD	International tourism, expenditures for passenger transport items (current US$)	International Trade	0.00022
BX.GSR.TOTL.CD	Exports of goods, services and income (BoP, current US$)	International Trade	0.00021
SH.XPD.PUBL	Health expenditure, public (% of total health expenditure)	Spending	0.00021
GC.XPN.OTHR.ZS	Other expense (% of expense)	Spending	0.00021
NE.CON.GOVT.CD	General government final consumption expenditure (current US$)	Spending	0.0002
BX.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, received (current US$)	Economic Environment	0.0002
SE.SEC.AGES	Secondary school starting age (years)	Workforce and Education	0.00019
SL.TLF.TOTL.FE.ZS	Labor force, female (% of total labor force)	Workforce and Education	0.00019
BX.GSR.GNFS.CD	Exports of goods and services (BoP, current US$)	International Trade	0.00019
NE.TRD.GNFS.ZS	Trade (% of GDP)	International Trade	0.00019
TG.VAL.TOTL.GD.ZS	Merchandise trade (% of GDP)	International Trade	0.00019

Table 19: Default occurrence estimation accuracy, restricted on most frequent underlying model accuracies

Variable Name	Variable Description	Theory Class	Importance Score in 2010 Model
NE.CON.PETC.KD	Household final consumption expenditure, etc. (constant 2000 US$)	Spending	0.00018
TX.VAL.MRCH.AL.ZS	Merchandise exports to economies in the Arab World (% of total merchandise exports)	International Trade	0.00018
TX.VAL.MRCH.R4.ZS	Merchandise exports to developing economies in Middle East and North Africa (% of total merchandise exports)	International Trade	0.00018
BX.TRF.CURR.CD	Current transfers, receipts (BoP, current US$)	International Trade	0.00017
ST.INT.TVLR.CD	International tourism, receipts for travel items (current US$)	International Trade	0.00017
TX.VAL.MMTL.ZS.UN	Ores and metals exports (% of merchandise exports)	International Trade	0.00017
ST.INT.DPRT	International tourism, number of departures	International Trade	0.00017
TM.VAL.SERV.CD.WT	Commercial service imports (current US$)	International Trade	0.00017
BN.TRF.CURR.CD	Net current transfers (BoP, current US$)	International Trade	0.00016
SE.SEC.ENRL.GC	Secondary education, general pupils	Workforce and Education	0.00016
ST.INT.RCPT.CD	International tourism, receipts (current US$)	International Trade	0.00016
TX.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service exports)	International Trade	0.00015
GC.TAX.GSRV.RV.ZS	Taxes on goods and services (% of revenue)	Tax	0.00015
BM.GSR.NFSV.CD	Service imports (BoP, current US$)	International Trade	0.00015
NE.CON.PRVT.KD	Household final consumption expenditure (constant 2000 US$)	Spending	0.00015
FS.AST.PRVT.GD.ZS	Domestic credit to private sector (% of GDP)	Economic Environment	0.00014
BM.GSR.GNFS.CD	Imports of goods and services (BoP, current US$)	International Trade	0.00014
IT.MLT.MAIN	Telephone lines	Economic Environment	0.00014
BM.TRF.PRVT.CD	Private current transfers, payments (BoP, current US$)	International Trade	0.00013
BX.PEF.TOTL.CD.WD	Portfolio equity, net inflows (BoP, current US$)	International Trade	0.00013
SL.TLF.CACT.ZS	Labor participation rate, total (% of total population ages 15+)	Workforce and Education	0.00013
BX.GSR.MRCH.CD	Goods exports (BoP, current US$)	International Trade	0.00013
MS.MIL.XPND.GD.ZS	Military expenditure (% of GDP)	Spending	0.00013
BM.GSR.TRVL.ZS	Travel services (% of service imports, BoP)	International Trade	0.00012
GC.XPN.GSRV.ZS	Goods and services expense (% of expense)	Spending	0.00011
SL.EMP.TOTL.SP.FE.ZS	Employment to population ratio, 15+, female (%)	Workforce and Education	0.00011
BX.GSR.NFSV.CD	Service exports (BoP, current US$)	International Trade	0.00009
ST.INT.RCPT.XP.ZS	International tourism, receipts (% of total exports)	International Trade	0.00009
SE.PRM.ENRR	School enrollment, primary (% gross)	Workforce and Education	0.00009
SL.TLF.TOTL.IN	Labor force, total	Workforce and Education	0.00009
TX.VAL.SERV.CD.WT	Commercial service exports (current US$)	International Trade	0.00008
NE.CON.TETC.KD	Final consumption expenditure, etc. (constant 2000 US$)	Spending	0.00007
BX.GSR.TRAN.ZS	Transport services (% of service exports, BoP)	International Trade	0.00007
BM.GSR.MRCH.CD	Goods imports (BoP, current US$)	International Trade	0.00007
NE.CON.GOVT.KD	General government final consumption expenditure (constant 2000 US$)	Spending	0.00007
BG.GSR.NFSV.GD.ZS	Trade in services (% of GDP)	International Trade	0.00006
BM.GSR.TOTL.CD	Imports of goods, services and income (BoP, current US$)	International Trade	0.00005
BX.GSR.CMCP.ZS	Communications, computer, etc. (% of service exports, BoP)	International Trade	0.00004
TM.VAL.MRCH.AL.ZS	Merchandise imports from economies in the Arab World (% of total merchandise imports)	International Trade	0.00004
ST.INT.ARVL	International tourism, number of arrivals	International Trade	0.00004
SE.PRM.ENRR.MA	School enrollment, primary, male (% gross)	Workforce and Education	0.00004
TM.VAL.TRVL.ZS.WT	Travel services (% of commercial service imports)	International Trade	0.00002
NE.CON.PRVT.PP.CD	Household final consumption expenditure, PPP (current international $)	Economic Environment	0.00002
SE.SEC.DURS	Secondary education, duration (years)	Workforce and Education	-0.00003
BN.GSR.FCTY.CD	Net income (BoP, current US$)	Economic Environment	-0.00003
SE.PRM.DURS	Primary education, duration (years)	Workforce and Education	-0.00007
BN.KAC.EOMS.CD	Net errors and omissions, adjusted (BoP, current US$)	International Trade	-0.0003